

Title: Genomic Considerations for FHIR; eMERGE Implementation Lessons

Authors: *Mullai Murugan¹, *Lawrence J. Babb³, Casey Overby Taylor⁴, Luke V. Rasmussen⁵, Robert R. Freimuth⁶, Eric Venner^{1,2}, Fei Yan¹, Victoria Yi¹, Stephen J. Granite⁴, Hana Zouk^{7,8}, Samuel J. Aronson⁹, Kevin Power¹⁰, Alex Fedotov¹¹, David R. Crosslin²¹, David Fasel¹², Gail P. Jarvik¹³, Hakon Hakonarson¹⁴, Hana Bangash¹⁵, Iftikhar J. Kullo¹⁵, John J. Connolly¹⁴, Jordan G. Nestor¹⁶, Pedro J. Caraballo¹⁷, WeiQi Wei¹⁸, Ken Wiley¹⁹, Heidi L. Rehm^{3,20}, Richard A. Gibbs^{1,2}

* Both authors contributed equally

1. Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX, USA
2. Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, USA
3. Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, USA
4. Departments of Medicine and Biomedical Engineering, The Johns Hopkins University School of Medicine, Baltimore, MD, USA
5. Department of Preventive Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL, USA
6. Department of Digital Health Sciences, Center for Individualized Medicine, Mayo Clinic, Rochester, MN, USA
7. Laboratory for Molecular Medicine, Mass General Brigham Personalized Medicine, Cambridge, MA, USA
8. Department of Pathology, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA
9. Partners Personalized Medicine, Partners HealthCare, Cambridge, MA, USA
10. Cerner Corporation, Kansas City, MO, USA
11. Irving Institute for Clinical and Translational Research, Columbia University, New York, NY, USA
12. Department of Medicine, Columbia University Irving Medical Center, New York, NY, USA

13. Department of Medicine (Medical Genetics) and Genome Sciences, University of Washington, Seattle, WA, USA
14. Center for Applied Genomics, Children's Hospital of Philadelphia, PA, USA
15. Department of Cardiovascular Medicine, Mayo Clinic, Rochester, MN
16. Department of Medicine, Division of Nephrology, Columbia University, New York, NY, USA
17. Department of Medicine and Department of Health Sciences Research, Mayo Clinic, Rochester, MN, USA
18. Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN, USA
19. National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA
20. Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA
21. Department of Biomedical Informatics and Medical Education, University of Washington, Seattle, WA, USA

Abstract

Structured representation of clinical genetic results is necessary for advancing precision medicine. The Electronic Medical Records and Genomics (eMERGE) Network's Phase III program initially used a commercially developed XML message format for standardized and structured representation of genetic results for electronic health record (EHR) integration. In a desire to move towards a standard representation, the network created a new standardized format based upon Health Level Seven Fast Healthcare Interoperability Resources (HL7 FHIR), to represent clinical genomics results. These new standards improve the utility of HL7 FHIR as an international healthcare interoperability standard for management of genetic data from patients. This work advances the establishment of standards that are being designed for broad adoption in the current health information technology landscape.

Keywords: HL7 FHIR Standards, Clinical Genomics, Genetic Test Results, Interoperability, Electronic health record (EHR), Clinical Decision Support

1. Introduction

The use of genomic testing in healthcare has grown in recent years, attributed to our increased understanding of the human genome and broader accessibility of tests[1–4]. Typically, the results from genetic testing laboratories are represented in a narrative, unstructured form, delivered as a Portable Document Format (PDF) document, which is simply scanned or uploaded to the electronic health record (EHR)[5]. This approach, though widely accepted and ubiquitous in its use, is primarily targeted to human readers. PDF and other custom structured reporting, though serving the immediate need of delivering genomic test results, hinders widespread reanalysis, data sharing, interoperability, automation, and downstream research efforts. Furthermore, without computational standards we are unable to reliably exchange genomic data and use those data to scale discovery and improve clinical care through automated clinical decision support (CDS)[6]. Because access to relevant genomic data in a standardized computable format has been a major hurdle to widespread implementation of genomic medicine, standards organizations, research networks and healthcare institutions have embarked on multiple projects in the past decade to improve the accessibility of genomic test results as computable data[7–9].

The Electronic Medical Records and Genomics (eMERGE) Network, funded by the National Human Genome Research Institute (NHGRI), is a consortium of U.S. medical research institutions that “develops, disseminates, and applies approaches to research that combine biorepositories with electronic medical record systems for genomic discovery and genomic medicine implementation research”[10]. Phase III of this program, as part of the Network’s overarching goal to integrate genomic test results to the EHR for clinical care, sought to create a proof of concept, standards-based representation of clinical genomic test results using the rapidly evolving Health Level Seven (HL7) Fast Healthcare Interoperability Resources (FHIR)[11] standard.

1.1. eMERGE Background

Currently in its fourth phase, the eMERGE Network’s first phase began in September 2007. eMERGE Phase III (September 2015 to March 2020) joined together multiple laboratories and clinics, including two central sequencing and genotyping facilities (CSGs): Baylor College of Medicine Human Genome Sequencing Center (BCM-HGSC) and the Broad Institute (BI) and Partners Laboratory for Molecular Medicine (LMM), as well as eleven study sites. The CSGs performed sequencing, variant interpretation, and clinical report generation. The reports were then returned to clinical sites and integrated into the respective EHRs[12,13]. As illustrated in Figure 1[14], the complexity and heterogeneity of Phase III of the Network provided considerable challenges for technical and data harmonization across the CSGs and study sites[15]. The network identified workflow and logistic components to harmonize both data and delivery of results, as well as the need for both a narrative PDF form to satisfy clinicians and regulatory requirements and a standardized XML format to represent clinical genetic test results directly in the EHR.

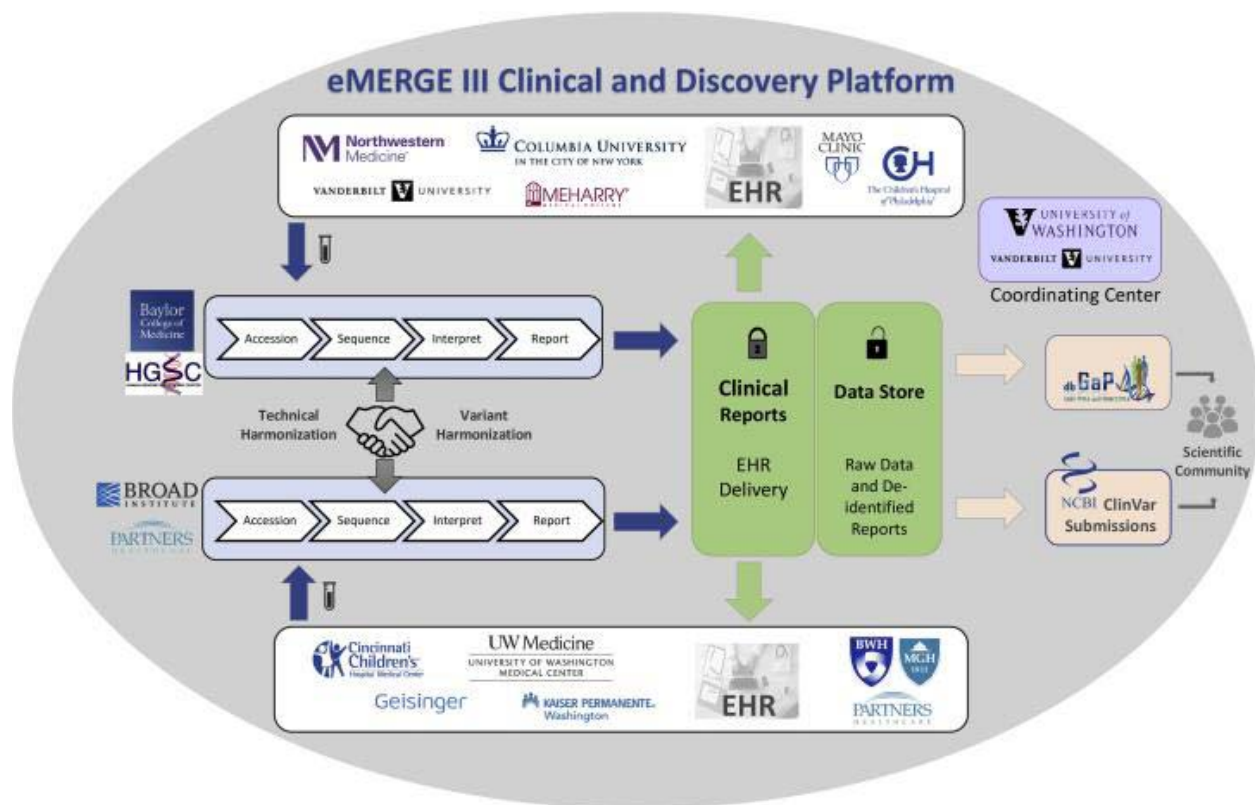


Figure 1. The eMERGE III Network. The Network comprised 11 study sites, 2 central sequencing and genotyping facilities (CSGs) and a coordinating center (CC) with samples sent from the study sites to the CSGs. Clinical Genetic Results were returned to the study sites after sequencing, variant classification and harmonization at the CSGs with raw data and de-identified results being returned to the CC for research. Both the clinical and research data flow is illustrated in this figure[14].

The electronic structure utilized by the eMERGE Network for genetic test reports included an XML format based primarily upon the features of Sunquest Mitogen™ Genetics (formerly known as GeneInsight), a commercial tool for clinical genetic test reporting and knowledge management originally developed by one of the CSGs[15]. This XML format enabled report data exchange between the eMERGE sites that used GeneInsight, the sites that employed local data, and the centralized repositories. The GeneInsight format, though opportune for the eMERGE Network at delivering the initial requirements, is not consistent with other open standards and thus not viable as the long-term solution required to support broader growth and adoption. With the emergence of FHIR as an interoperable healthcare standard and the work of the HL7 Clinical Genomics Workgroup (CG WG) towards the creation of a FHIR Genomics

Reporting Implementation Guide (GR IG)[16], the Network decided on a specification based on the GR IG in an effort to contribute to and validate the nascent GR IG.

1.2. FHIR Background

The FHIR standard builds on the widely used HL7 Version 2[17] messaging standard and leverages common, modern web technologies that facilitate the development and integration of applications into clinical environments. FHIR has gained traction since its inception in 2012 as an open API standard for interoperability, continuing to grow with programs such as the *FHIR Accelerator Program*[18], the healthcare pledge by technology companies to adopt emerging standards such as FHIR for interoperability[19], along with NIH notices encouraging the use of FHIR for clinical and research interoperability[20,21]. Clinical software vendors are also increasingly implementing support for FHIR-based APIs into EHRs and laboratory information management systems (LIMS), enabling clinical lab test results to be rendered as discrete data elements that can be used to drive CDS.

Building upon the core FHIR standard, the FHIR community has worked to support its implementation for clinical genetics and genomics. In November 2019, the HL7 CG WG published the first release of the GR IG, which is intended to support "all aspects[22]" of reporting clinical genomics data. The GR IG contains specific guidance for reporting genomic variants that pertain to pharmacogenomics (PGx), tumor testing, and histocompatibility profiling. The current release of the GR IG [16], based on FHIR R4[22], has been balloted as a Standard for Trial Use (STU), which is intended to be vetted through implementation and pilot testing by early adopters. This process is critically important to the advancement of the standard as feedback drives improvements in the specification before it becomes more widely adopted, which limits the ability to make breaking changes as gaps are filled.

1.3. Project Objectives

In 2019 the eMERGE Network aimed to improve structured data standards and take advantage of newly emerging FHIR capabilities. Three milestones were established: 1. Develop a computable and standardized clinical reporting specification for eMERGE with FHIR; 2. Create a proof-of-concept implementation pilot, generating eMERGE Clinical Genetic Reports with the eMERGE FHIR Specification; and 3. Establish EHR interoperability with FHIR-enabled secure ingestion of genetic results and implement clinical decision support use cases.

Recognizing the emerging nature of this landscape, the network limited the scope of its effort to building a baseline upon which future work can be continued. Here we describe these efforts, focusing on identifying lessons learned and future considerations on how organizations can begin to adopt the FHIR genomics standard, illustrating the level of expertise and effort required to do so, and demonstrating the positive impact that such efforts can have on the development of international standards.

2. Methods

Fulfillment of the objectives and the milestones set forth for this project was managed as a cross-network collaboration with the two CSGs first developing the eMERGE FHIR Specification. The BCM-HGSC acted in its role of a diagnostic laboratory to generate a sample set of clinical genetic reports utilizing the specification, and study site Northwestern University (NU) and non-clinical affiliate Johns Hopkins University (JHU) acted as provider facilities that implemented a proof-of-concept EHR integration and clinical care pilot to demonstrate the viability of the eMERGE FHIR Specification. These steps are illustrated in Figure 2 and detailed further in this section.

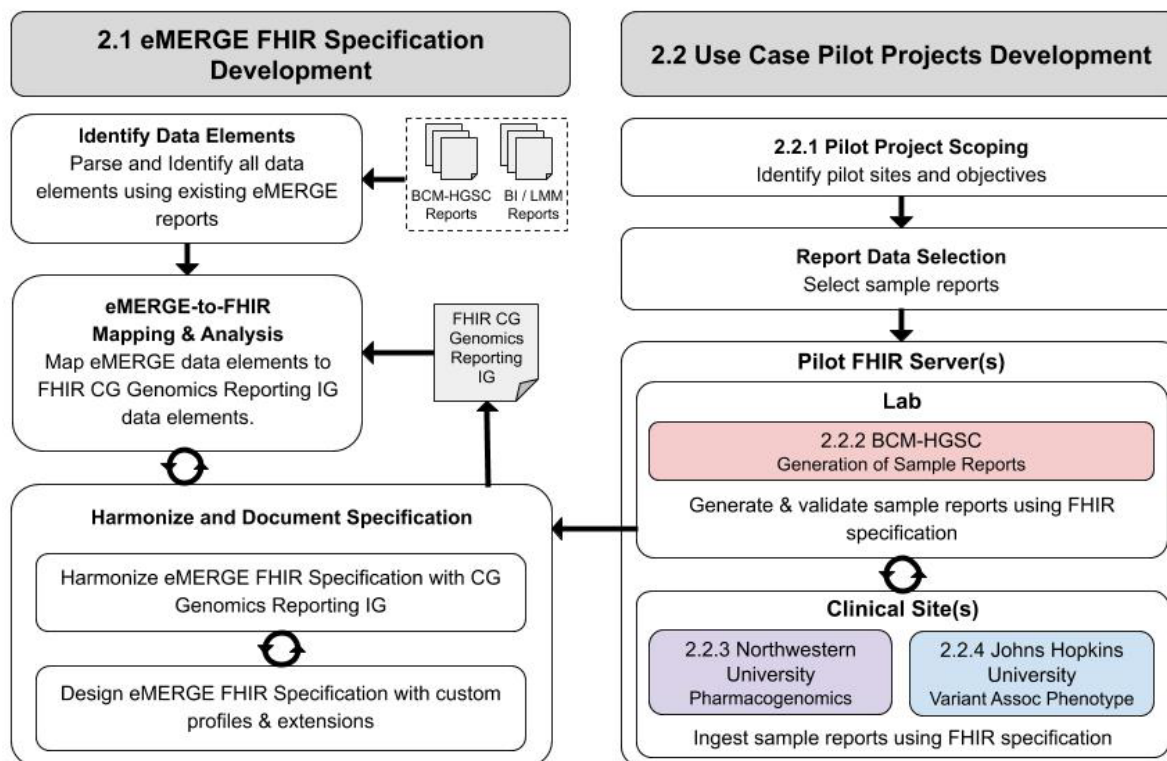


Figure 2. Methods for specification and pilot projects development. The two columns represent complementary work streams - the leftmost dedicated to the specification development, and the right on use cases for pilot projects development. The solid arrows represent the stepwise or iterative direction of the respective processes.

2.1. eMERGE FHIR Specification Development

The two CSGs established the following guiding principles for this project, taking advantage of their respective experiences:

- **Structured content** - All content from the narrative PDF eMERGE reports and all eMERGE standard reporting use cases should be captured in structured format and as meaningful data elements without losing content and context;
- **Alignment with HL7 FHIR Core and GR IG** - All eMERGE concepts and associated elements shall be aligned with GR IG and FHIR Core Standards and extended as required;
- **Computationally reliable representation of results** - An optimal computational form for each data element shall be determined, prioritizing eMERGE pilot objectives for EHR

integration and CDS. All specimen types and genetic data elements related to the resulting observations must be based on reference sequences, coordinates and structures that consistently and accurately reflect the lab methods used to align the raw data, determine coverage and call the variants; and

- **Codify concepts when reasonable** - Concepts should be codified using FHIR Core and GR IG guidance. eMERGE concepts that extend beyond the FHIR and CG guidance should be codified if possible and within reason.

The development of the eMERGE FHIR Specification consisted of the following steps -

- **Identify data elements** using existing results from a comprehensive set of eMERGE reporting use cases;
- **Map eMERGE report** elements and structures both semantically and structurally to GR IG resources and profiles; perform an analysis, identify issues that require further resolution, and propose resolutions; and
- **Harmonize and document** finalized decisions informed by
 - Harmonizing changes with GR IG
 - Documenting resolutions requiring custom profiles & extensions
 - Feedback from BCM-HGSC lab pilot development

The steps defined above were accomplished by first analyzing eMERGE narrative reports representing a range of use cases. Next, working with domain experts (geneticists from both the CSGs), the semantic meaning and structure of these concepts were verified and all data elements appearing on the reports were associated with the proper concepts. This step was critical to assuring that at a minimum both CSGs agreed with the conceptual meaning of all granular elements on their reports, minimizing subsequent misrepresentation of the elements.

The next step was to determine how best to computationally represent these concepts. A gap analysis was performed between each eMERGE concept and its corresponding GR IG resources and profiles. The aim of this effort was to synchronize GR IG resource profiles to related eMERGE reporting concepts. Gaps, questions and issues arising during this process were publicly documented and harmonized between the eMERGE CSGs and the HL7 FHIR community. This communication consisted of informal feedback through the HL7 FHIR Zulip Chat Board[23], and formal feedback through issue trackers on Jira[24]. Complex questions were addressed through discussions with the Clinical Genomics WG or the eMERGE EHRI

Workgroup. This reconciliation and harmonization process resulted in a beta version of the eMERGE FHIR Specification. This process also contributed to enhancements and changes to the GR IG itself. Implementation of the use case pilot projects contributed to the validation of the eMERGE FHIR Specification and generated additional feedback that was integrated into the specification.

2.2. Use Case Pilot Projects Development

2.2.1. Pilot Project Scoping

We conducted a proof-of-concept implementation pilot that represented real clinical genetic testing reports using the eMERGE FHIR Specification in EHRs. The pilot project included three use cases: 1. BCM-HGSC - Generation of sample reports using the eMERGE FHIR Specification, 2. NU - Pharmacogenomics clinical decision support, and 3. JHU - Variant associated phenotype. The pilot further allowed ratification of the eMERGE FHIR Specification via iterative feedback from the test sites.

2.2.2. BCM-HGSC - Generation of Sample Reports

The pilot included representative data from a wide cross section of samples, including positive and negative results, heterogeneous indications for testing, multiple variant and PGx findings, selected phenotype susceptibilities, sex, and race. Reports for the sample cohort were generated in the eMERGE FHIR Specification.

The BCM-HGSC developed a custom FHIR Client[25] using the HAPI FHIR API[26] to convert the raw JSON results output of the BCM-HGSC's Clinical Pipeline Neptune to FHIR Bundles consisting of resources, profiles, extensions and coding systems as identified in the specification. A secure HIPAA compliant Amazon Web Services (AWS) cloud-based SMILE CDR[27] environment was established with the necessary GR IG profiles, FHIR extensions, and relevant coding systems such as Logical Observation Identifiers Names and Codes (LOINC), Systematized Nomenclature of Medicine (SNOMED), and the HUGO Gene Nomenclature Committee (HGNC). This environment was used to validate the structure and content of the FHIR Bundles. The FHIR Bundles themselves were stored on the SMILE CDR server and on AWS S3.

During this implementation, the group iteratively revised and refined the specification based on the implementation's efficacy assessment, and validated results with FHIR Core and GR IG to ensure adherence. Finally, the BCM-HGSC tested and internally validated FHIR bundles, representing the sample cohort, for data integrity and accuracy using the narrative PDF. The content of the FHIR reports such as patient/sample identifiers and results were validated against the source JSON via automation. A set of three FHIR reports were manually reviewed end to end for structure, format and content by the developer (F.Y.). Once validated, FHIR bundles were distributed to JHU and NU.

2.2.3. Northwestern University - Pharmacogenomics

The NU use case targeted integrating FHIR results into an existing Ancillary Genomics System[28] (AGS), which was used as an intermediary platform to transmit results to the EHR. Pre-existing PGx CDS alerts were configured within the EHR to alert providers to recommendations for clopidogrel, warfarin, and simvastatin, and are available from the CDS Knowledgebase[29].

In order to complete this use case, NU required access to the PGx interpretations within the overall report, including genes, diplotypes, and medications. Because of institutional constraints surrounding the use case, the system was not connected to the live EHR environment. Instead, given that a validated workflow existed for the same data, we considered success as the ability to import the FHIR bundles into a test FHIR server, and the ability to transform the FHIR resources into the same intermediate results that were sent to the EHR to trigger CDS.

NU implemented a locally hosted instance of the open-source Microsoft FHIR Server[30], connected to a Microsoft SQL Server 2017 Docker container as the underlying data store, and configured with the FHIR R4 Specification. A command line application was created in C# using the .NET Core framework, and is freely available on GitHub[31]. This utility loaded FHIR bundles stored on disk to the FHIR server in order to simulate the process of "receiving" results from an external laboratory. In addition, the utility then extracted data from the FHIR Server to the simulated AGS. The import process was manually validated for a set of 25 randomly selected patients by one of the authors (L.V.R) to confirm the correct attribution of diplotypes to patients.

2.2.4. Johns Hopkins University - Variant Associated Phenotype

The goal of the JHU use case was to document in the EHR the phenotype(s) known to be associated with a pathogenic or likely pathogenic variant and the corresponding evidence provided by the diagnostic lab in the FHIR report. A cloud based FHIR service for FHIR R4 and DSTU3 was deployed as an AGS to provide EHR access to genomic lab report data. In addition, a secure FHIR Genomics Proxy instance was deployed[32] with an authorization access role for BCM-HGSC users to send FHIR bundles to the AGS. With proper authentication, JHU accepted FHIR-formatted reports from BCM-HGSC and stored those reports on the cloud-based FHIR server. JHU used the Azure API for FHIR[33] and registered it in the JHU API portal[34] so that it could be used to add a “variant associated phenotype” to the patient chart for a set of test patients using a web service supported by Epic(R)[35].

In order to validate the JHU import process, a BCM-HGSC application was generated to create FHIR transaction bundles and submit them to the JHU FHIR server, while the JHU team generated a Validator tool. The number of bundles submitted by BCM-HGSC and OMIM identifiers in those reports found in the JHU AGS were reported. (See **Appendix A** Johns Hopkins University Technical Validation).

2.2.5. Assessment of Data Quality for Use Cases

For each use case we assessed data quality characteristics of the electronic patient data used by the NU and JHU sites for their CDS use cases. Our approach adapted a data quality scoring system to score the FHIR transaction bundles according to four characteristics (Completeness, Accessibility, Relevance/Fitness, and Reliability/Accuracy). **Appendix B** provides details about the methods to assess data quality that NU and JHU applied after completing their use cases.

3. Results

3.1. eMERGE FHIR Specification Development

The principal outcomes of the specification development were the following: 1. To identify the complete set of report concepts and elements used throughout all eMERGE reporting use cases; 2. To create a FHIR based schema using the GR IG that would be implementable by

eMERGE; 3. To provide a public document of the eMERGE FHIR Specification; and 4. To collaborate with the HL7 CG WG to harmonize useful feedback into the GR IG and FHIR Specification in general.

3.1.1. Identification of eMERGE Report Concepts and Elements

The first step towards the creation of the eMERGE FHIR Specification was an “As Is” analysis of the existing genetic reports to inventory all eMERGE reporting concepts and elements. To this end, we compiled a set of all-inclusive representative reports from both the CSGs (see Figure 3 for an example report from each CSG) to ensure use cases requiring unique report concepts and elements were included. Table 1 contains the list of principal eMERGE reporting scenarios.

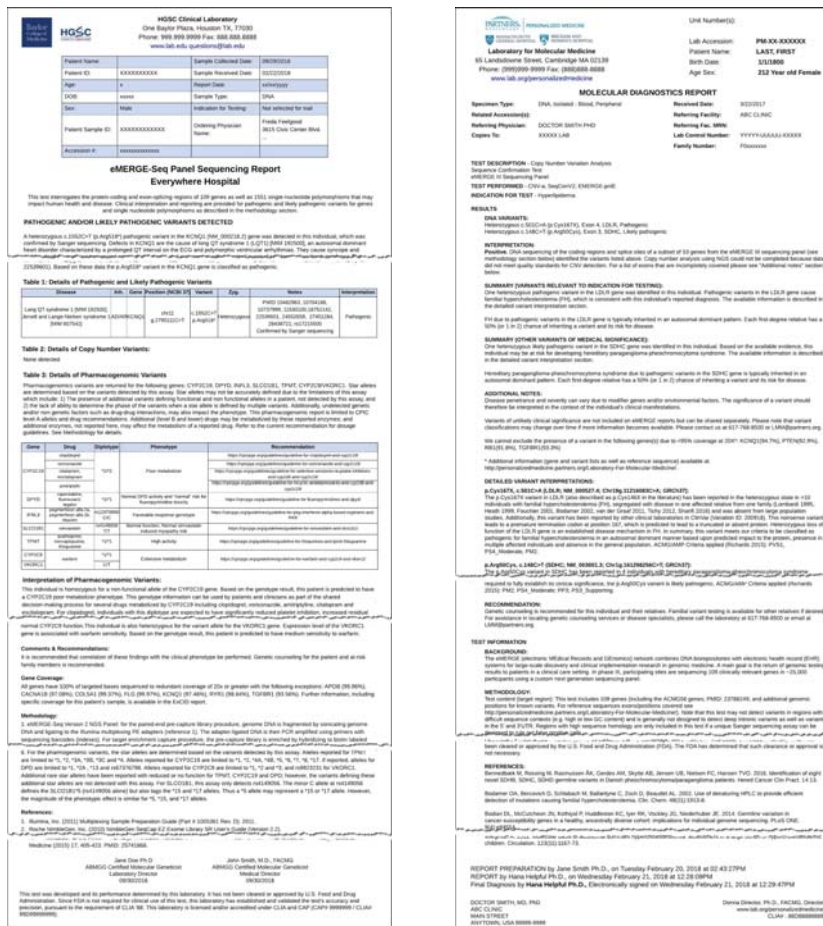
Standard Reporting Use Cases	
Diagnostic Gene Panel	
Reported Variants	Single-nucleotide variants (SNVs), Insertions or deletions (InDels), copy number variants (CNVs)
Medically Significant Findings for Indication	Pathogenic, Likely Pathogenic, Uncertain Significance
Secondary Findings	Medically significant findings in a gene unrelated to the primary indication for testing
Overall Interpretations	Positive, Negative, Inconclusive
Pharmacogenomic Gene Panel	
PGx Phenotype Findings	Efficacy, Metabolism, Transporter Function
Reported Diplotypes	Simple, Complex, Wildtype, Non-Wild Type
Specialized Reporting Use Cases	
A polygenic score for LDL cholesterol	Mayo Clinic, an eMERGE Study Site
Research Only Reports	Columbia University, an eMERGE Study Site

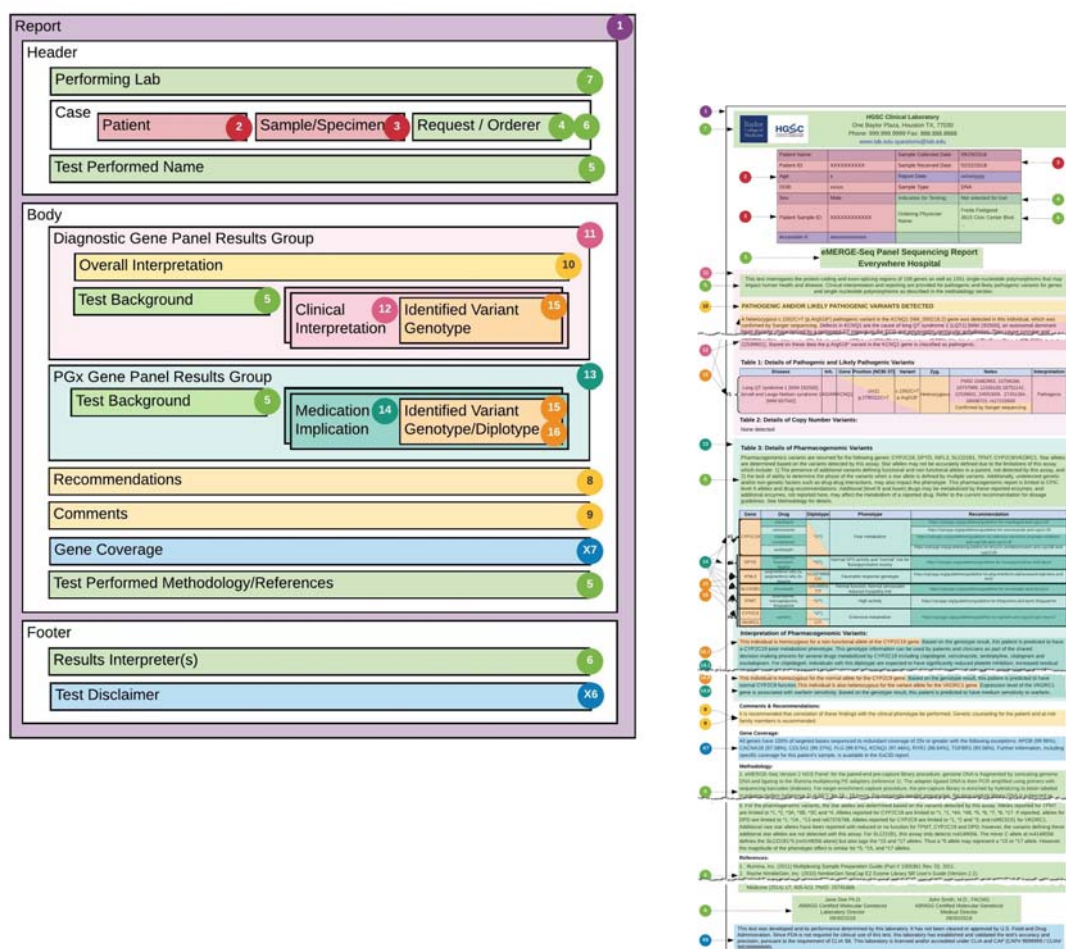
Table 1: eMERGE Reporting Use Cases. This table lists all of the eMERGE reporting use cases. The first column is a breakdown of reporting result types and the second column lists more specific outcomes for the corresponding reporting result type. The specialized reporting use cases are specified for the associated eMERGE study sites.

Using selected reports for these associated use cases, the structure and composition of the reports was analyzed, and a set of data elements was assembled (Figure 4), resulting in 18

core concepts and around 100 fundamental data elements. This analysis and documentation of the existing eMERGE report content served as the foundation for the design of the eMERGE FHIR Specification.

In an effort to manage the pilot within allotted resources and time, the scope of this effort was confined to the Standard Reporting Use Cases and result delivery, while including provisions for future expansion.





major eMERGE report concepts were aligned to the GR IG resources and profiles, followed by a granular mapping of every eMERGE report element to a corresponding FHIR resource element.

3.1.2.1. Mapping eMERGE Report Concepts to FHIR Resources and Profiles

The GR IG provided the guidance for driving the mapping of the eMERGE report concepts to its resources, profiles and extensions. Our first attempt at mapping resulted in several key structural and organizational questions; as these questions and issues were fundamental to the design of the eMERGE FHIR Specification, a significant amount of time was spent to identify long and short-term resolutions.

The end result was a set of 21 noteworthy issues and associated resolutions, documented as part of the eMERGE Specification[36]. Below is a representative issue and its resolution followed by a complete listing of all 21 issues that were addressed (Table 2). These issues and their resolutions vary in scope and complexity with proposed solutions being driven by balancing the timeline for delivering the eMERGE pilot against the ideal solution as noted.

#1 Composite reporting[37]

Diagnostic labs frequently include several types of results in genetic test reports which then appear in separate sections within an overarching report. eMERGE genetic reports are examples of this model and include both gene panel interpretation as well as PGx results. This style of reporting is analogous to the notion of composite reporting whereby while individual sections of the report can be treated independently, they are still combined together as they rely on shared findings from an upstream wet lab assay such as Whole Exome Sequencing or Gene Panels. We evaluated two options to represent composite reporting: 1. Nested Diagnostic Report Resources or 2. The Observation Grouper Profile[38]. Based on analysis and collaborative discussions with the CG WG[39], we decided on the Grouper Profile.

The resolution to use the Observation Grouper Profile allows consuming EHR systems to utilize the Gene Panel & PGx results as disparate components. Furthermore, while the usage of Grouper Profile was well-defined by the GR IG, the concept of nested Diagnostic Reports was still under investigation and not ready for adoption. The benefit of this approach also lends itself to including additional sections such as Polygenic Risk Scores to genetic reports.

Issue	Cat.	Resolution	Issue	Cat.	Resolution
-------	------	------------	-------	------	------------

#1 Composite Reporting	Major	GR IG Change	#12 Pathogenicity Values	Minor	GR IG Change
#2 Nested Results	Major	Clarification	#13 Report Category	Minor	GR IG Change
#3 Test Information	Major	Clarification	#14 Assessed Med Citations	Major	Workaround
#4 Interpretation Summary	Major	Workaround	#15 Sign-Out v Sent Date	Minor	Workaround
#5 Gene/Region Coverage	Major	Workaround	#16 Recommended Followup Reasons	Minor	Workaround
#6 Recommendations	Major	Clarification	#17 Recommended Followup Codes	Minor	GR IG Change
#7 Secondary Findings	Major	Clarification	#18 Disclaimers	Major	Workaround
#8 Variant Data Types	Major	Deferred	#19 Patient Internal ID	Minor	Clarification
#9 Report Comments	Minor	Clarification	#20 Patient Age	Major	Workaround
#10 Confirmation Testing	Major	Workaround	#21 Research Flag	Minor	Deferred
#11 Pathogenicity Phenotypes	Minor	GR IG Change			

Table 2: Issues & Resolutions. This table lists all of the noteworthy issues that required investigation and collaboration with the HL7 CG WG to resolve. The three columns show the issue label, category (*Major*, *Minor*) and resolution status (*Clarification* - GR IG or FHIR solution, *STU1 Change* - CG WG modified GR IG to support, *Workaround* - extension or non-standard approach, *Deferred* - deferred workaround but submitted for future consideration). Detailed description for each of these issues is available on the Issues & Resolutions[36] page of the eMERGE Read the Docs website[40].

Addressing and resolving these issues resulted in the mapping and structural design of the specification, illustrated in Figure 5. As illustrated, the root profile of the specification is the GenomicsReport; this is the key resource that encapsulates the ServiceRequest for the test, the Observations that constitute the results (i.e. findings or implications of the test), the Tasks that include clinical care recommendations, and the Grouper Profile to organize and manage composite resulting (i.e. GenePanel and PGx results). Other major resources attached to the GenomicsReport include the Patient for whom the test is being ordered, the associated Specimen, the Practitioner ordering the test, the Organization (i.e. Diagnostic Laboratory performing the test) and the Practitioner interpreting the results of the test.

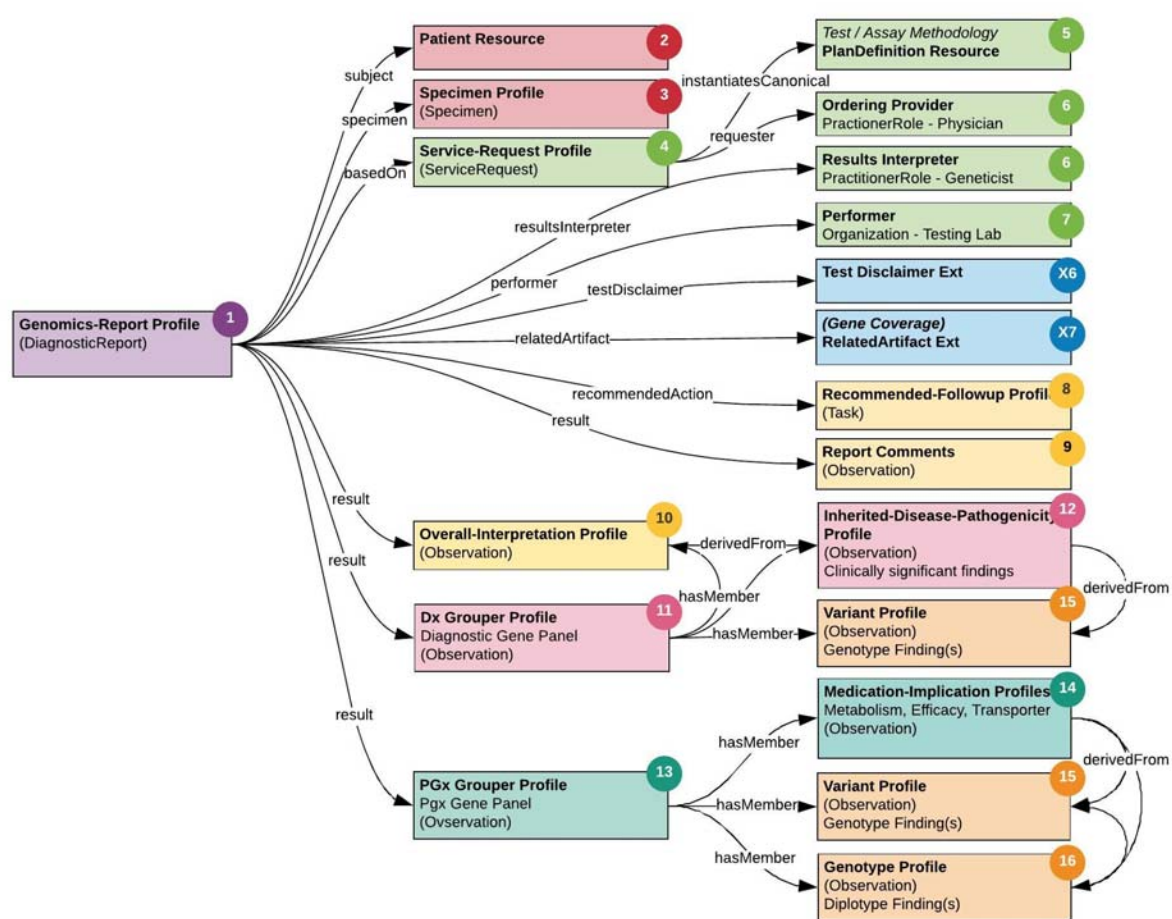


Figure 5. Structural Design of the eMERGE FHIR Specification. A color-coded organizational mapping of the FHIR Genomics Reporting Implementation Guide resources and profiles for eMERGE, including custom eMERGE extensions in blue (X6, X7). The color and numbering maps back to the report layout and example in Figure 4 for easier reference.

3.1.2.2. Mapping eMERGE Field Level Report Elements to FHIR Resource Elements

Step 2 involved a granular mapping of every eMERGE report attribute to an equivalent field in the FHIR resources identified in the previous step. This was a laborious process which in addition to requiring precise and careful mapping of the fields themselves, also required determining naming systems and assignment of coding systems, codes and values. Online documentation[41] includes the complete set of eMERGE FHIR resources and its associated elements, with a summary listed in Table 3. Furthermore, gap analysis at this step revealed the need for additional fields such as summary interpretation text, test disclaimer etc. that were not available in the GR IG. Though we documented these as feature requests in HL7's Jira[24], to

satisfy the immediate needs of the project, we created these fields as FHIR Extensions. The full list of Extensions[42] is available on the eMERGE Read the Docs website[40].

Resources	Profile or Resource	Elements				
		Total	Ref	Ext	Id/Code	Other
GenomicsReport	profile	22	9	3	6	4
Report Resources						
Patient	resource	10	0	4	3	3
Specimen	profile	11	2	0	5	4
RecommendedFollowup	profile	10	2	0	5	3
ServiceRequest	profile	17	5	0	9	3
PlanDefinition	resource	11	0	0	4	7
Observation Resources						
OverallInterpretation	profile	15	5	1	6	3
ReportComment	resource	12	4	0	5	3
GrouperDx	profile	16	7	1	5	3
GrouperPgx	profile	17	9	0	5	3
InheritedDiseasePathogenicity	profile	19	5	1	9	4
MedicationImplication*	profile	17	6	1	7	3
Variant	profile	31	4	0	21	6
Genotype	profile	15	5	0	7	3
Other Resources						
PractitionerRole	resource	6	2	0	3	1
Practitioner	resource	7	0	0	4	3
Organization	resource	9	1	0	4	4
Total Concept Elements		223	63	11	97	52

Table 3: eMERGE FHIR resources. This table lists all resources and profiles in the eMERGE FHIR report bundle, with counts of the total number of elements (Total Elements) in each resource. Also included are a breakdown of the counts for each type of element i.e. references to other resources within a resource (Ref Elements), extensions used and/or created specifically for eMERGE (Ext Elements), naming or coding Systems (Id/Code Elements) and others (Other Elements).

3.1.3. Harmonization and Documentation

The eMERGE FHIR Specification project was initiated while the HL7 CG WG was still working on its first draft of GR IG. Though the draft was a good start for the eMERGE Specification, it did not encompass all the eMERGE use cases. Additional documentation for the eMERGE FHIR Specification is available on the eMERGE Read the Docs website[40]; the sample reports generated via an implementation of the specification are available on GitHub[43].

3.2. Use Case Pilot Projects Development

The principal goals of the pilot were to a) demonstrate the successful generation of sample reports adhering to the format of the eMERGE FHIR Specification, b) successfully ingest these reports to FHIR servers at the recipient sites, and c) demonstrate use of the FHIR formatted data for NU and JHU's respective use cases. Figure 6 illustrates the software components and data flow and is described in more detail in the following sections.

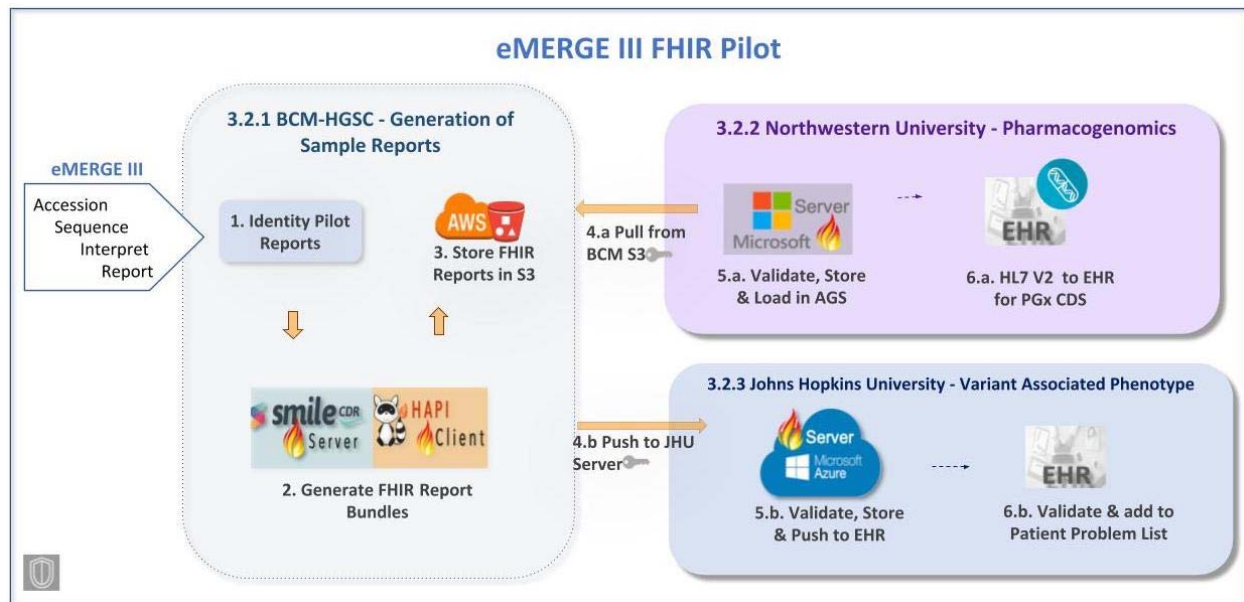


Figure 6: eMERGE III FHIR Pilot. FHIR bundles for the eMERGE pilot samples were generated by BCM-HGSC and pushed to NU and JHU's FHIR servers. Data was extracted from these FHIR reports for the NU PGx and JHU Variant Associated Phenotype use cases respectively.

3.2.1. BCM-HGSC - Generation of Sample Reports

3.2.1.1. Pilot Sample Selection

In order to facilitate the use cases for the pilot implementation, we selected 194 sample reports from the NU eMERGE study site (161 positive and 33 negative), and included a cross section of positives, negatives, heterogeneous indications for testing, multiple variants and PGx findings, selected phenotype susceptibilities, sex and race.

Taking into account JHU's role as a non-clinical eMERGE affiliate (meaning no genotyped data for JHU patients were available), we utilized the NU pilot set for JHU after de-identification. De-identification was performed within BCM-HGSC's HIPAA compliant environment on DNAnexus, and according to HIPAA Safe Harbor[44]. All identifiers were stripped of their original values and left empty, with the exception of patient and sample identifiers which were replaced with randomly generated UUIDs. This follows the data security and sharing guidelines established by eMERGE Consortium's Data Use Agreement (DUA)[45], to which all three parties involved with the pilot have agreed to comply for the pilot implementation. JHU, in their HIPAA compliant test environment, mapped these random UUIDs to patient MRNs, after matching on age, sex, and variant associated phenotypes reported, thereby creating a comparable equivalency for testing purposes.

3.2.1.2. Report Generation

A total of 194 reports were successfully generated and represented as FHIR bundles. Two copies of the reports were created - one containing the original identifiers for NU, and one containing de-identified reports for JHU. No issues were found in the validation of the reports done by the study team before transmitting to NU and JHU.

A set of five report bundles, two negatives and three positives, generated with this implementation and employing the eMERGE FHIR Specification are available for download on GitHub[43]. These sample reports were generated with simulated data being used for identifiers, namespaces and results.

3.1.2.3. Report Delivery

Dual modes of report delivery were employed: NU's FHIR bundles were downloaded directly from the BCM-HGSC provisioned AWS S3 and uploaded to an open-source Microsoft FHIR Server. JHU's FHIR bundles were pushed directly by the BCM-HGSC's FHIRClient to an Azure

Cloud Microsoft FHIR Server after successful authentication and authorization. Methods to integrate FHIR servers into the EHR at the pilot sites were considered out of scope for this paper.

3.2.2. Northwestern University - Pharmacogenomics

All of the 194 FHIR bundles selected for the pilot test were loaded to the FHIR server without error. The loading process to transmit the FHIR data into the simulated AGS resulted in the creation of 194 distinct patient records, and 1358 total diplotype records (7 diplotypes per patient). All data elements needed to trigger pre-existing CDS were present in the FHIR bundles. The manual review of 25 randomly selected patients confirmed that all diplotypes were loaded without error and attributed to the correct patient. Although no patient data was transmitted to the EHR, the process resulted in the exact same data available in the NU production environment, which is known to be sufficient for triggering CDS.

3.2.3. Johns Hopkins University - Variant Associated Phenotype

All of the 194 FHIR bundles selected for the pilot test were loaded to the JHU AGS without error, and the custom Validator tool verified that all 194 patients, their corresponding Specimens and all other attributes submitted by BCM-HGSC were identified (see **Figure A.1** in **Appendix A**). Processing the variant associated phenotype portion of the FHIR data captured in the JHU pilot AGS using the Validator tool found no discrepancies with what was expected (see **Figure A.2** in **Appendix A**).

3.2.4. Results from Assessing Data Quality for Use Cases

Findings from analyzing data quality characteristics of the eMERGE FHIR genetic report data for the NU and JHU CDS use cases indicated high ratings for all characteristics (Completeness, Accessibility, Reliability/Accuracy, and Relevance/Fitness) with one exception (see **Table B.2** in **Appendix B**). The *Completion* characteristic was rated lower for the JHU use case due to a need for additional processing of the report data in order to add a variant associated phenotype to a patient problem list using the FHIR API. Results are described in more detail in **Appendix B**.

4. Discussion

Through this eMERGE pilot effort we developed a standards-based specification for representing clinical genetic and genomic test results and demonstrated its viability with EHR integration and clinical use cases. This project supported the development of the FHIR Specification for genomics by basing the eMERGE FHIR Specification on the GR IG, and subsequently demonstrating use of the FHIR standard for genomic medicine. Through a data quality assessment framework (**Appendix B**), we confirmed that the data structures of the FHIR-formatted reports were acceptable for the two CDS use cases. Additionally, the deployment provided several lessons that guide future evolution of the FHIR Specification.

The CG WG released the first standard trial use (STU1) of the GR IG in November 2019, nearly a year after the work described in this manuscript began. As an early adopter, we worked collaboratively with the CG WG to further inform the design of the STU1 specification as it matured. As our experience demonstrated, feedback from early adopters contributes significantly to the development of a draft specification, but adopters must be flexible as the standard evolves in parallel.

The level of detail in the GR IG, which uses a series of FHIR profiles designed to reflect the complexity of clinical genomic test results, requires adopters to have an in-depth understanding of the domain as well as time and resources to implement the complex specification itself. This could be a deterrent to adoption for vendors, laboratories, or consumers that have relatively simple data.

The CG WG, recognizing that the complexity of the GR IG could be a barrier to adoption, is actively working with stakeholders to refine the specification. The experience of this project demonstrates the potential value of working directly with adopters to identify minimum viable products that are based on local use cases. Reciprocally, the CG WG would benefit significantly from concerted efforts by potential adopters (i.e. vendors, laboratories, healthcare systems, health IT engineers) through participation and collaboration to help inform the roadmap and development of the GR IG. Examples of such initiatives include ONC Sync for Genes[46], PGx CDS with FHIR and CDS Hooks[47], HLA Reporting with FHIR[48], Variant Interpretation for the Cancer Consortium[49], and mCODE[50].

4.4. Future Considerations

The issues we addressed during this project varied in scale in both importance as well as complexity. From these we have selected four key possibilities for further development: variant data types, interpretation summary text, gene/region coverage, and PGx results representation. A comprehensive description of all the issues can be found in the eMERGE FHIR Specification[36].

4.4.1. Variant Data Types

Community standards are required to define variants in a manner that will allow science and medicine to adopt and develop tools and procedures to track and attach information to variants at the scale demanded by genetic testing. Efforts to describe and identify variants include: Human Genome Variation Society (HGVS) nomenclature[51], International System for Human Cytogenetic Nomenclature (ISCN) nomenclature[52], and Variant Call File (VCF) formats[53]. Public repositories and registries established by various authorities have evolved over the past decade or more to provide identifiers for variants in these systems (e.g. dbSNP[54], ClinVar[55], ClinGen Allele Registry[56], COSMIC[57]).

The notion of a definitional or reference variant does not exist in FHIR or the GR IG, with FHIR and the CG WG lacking the data types and resources needed to build these referenceable variants. The CG WG's proposed resolution is to develop FHIR profiles based on the Observation resource. This begins to address the concern of separating the variant observation from the other assertion profiles, like disease causality or therapeutic implications of metabolism or efficacy. However, it does not provide a pragmatic computational standard for representing the variants referenced within those observed variants and associated assertions.

The Genomic Knowledge Standards (GKS) Workstream of the Global Alliance for Genomic Health (GA4GH) is developing and expanding the Variation Representation Specification[58,59] (VRS) to address the need for standards for computationally sharing variation. Instituting such a model in FHIR will significantly reduce the adoption risks caused by the complexity and unguided extensibility of the current GR IG and FHIR Specifications. As such, the growing collaboration between the CG WG and the GA4GH GKS Workstream represents a promising step forward at introducing the concepts, resources and data types needed in the FHIR Specification to improve the viability of implementing use cases related to variation in FHIR systems.

4.4.2. Interpretation Summary Text

While structured and coded results are of great importance to the computational utility of results, unstructured text will always play a significant role in conveying information between humans. Though there are a number of text attributes available throughout the GR IG resources, the genetics community requires the ability to associate an interpretation summary with the overall results of the report and with one or more classified variants. It is our recommendation that the CG WG consider all of the important kinds of text fields needed to support clinical genetic test results and assure that there is a mechanism to do so, starting with an interpretation summary text field.

4.4.3. Gene / Region Coverage

Because testing methods vary, it is important to provide a quantitative representation of the precise molecular sequenced regions covered and the quality of coverage for each region. Perhaps more importantly, this information clearly identifies which regions of the genome were not covered but may be relevant for the indication for testing. Receiving systems would then be able to accurately determine whether a patient may need follow-up testing to interrogate additional regions, and results across cohorts and studies would be more comparable.

4.4.4. PGx Results Representation

Key challenges to computationally representing PGx results are 1) to include the assayed variants that underlie the derived haplotype/diplotype assertions and 2) to maintain a clear distinction between the case-independent PGx knowledge statements about the haplotype/diplotype states and the patient-specific PGx implications expressed in the interpretation, which are based on the assay findings. The GR IG should provide a mechanism for defining variants at various levels that can be referenced either in the context of a patient observation or independently in a generalizable knowledge base. Also, it is currently difficult to convey the underlying assayed variants using the existing specification; if this could be accomplished then clinical systems could identify previously reported PGx results that may be impacted by future knowledge updates. This design approach supports the need for variant data types as discussed in section 4.4.1.

5. Conclusion

With this project, we have developed and demonstrated the use of a structured, computable and interoperable HL7 FHIR standard for eMERGE Clinical Reporting and have demonstrated the feasibility of implementing a standardized representation of clinical genetic and genomic results in the EHR. We have helped move the needle forward for the HL7 Clinical Genomics Workgroup in its development of the Genomics Reporting Implementation Guide by identifying and contributing to several growth areas[36] in order to effectively fulfill the return of clinical genomics results.

The two pilot use cases helped us understand the landscape to use FHIR-formatted reports for genomic medicine and its challenges. Establishing a standard, though fundamental, is but one step towards the effective use of genomic medicine in the clinic. In order to achieve this vision, additional work is needed not only to mature and stabilize a standard for structured representation of clinical genomic data but also with establishing engagement and collaboration between domain experts. Myriad perspectives are needed, including that of clinicians and geneticists, standards bodies such as HL7, EHR and other clinical system vendors, diagnostic laboratories, and health IT. These perspectives will further identify genomic medicine use cases and build standards and systems to support these use cases as well as target integration and interoperability amongst users, generators and consumers of clinical genomic data. Tenacious continuation of such undertakings and joint complementary efforts is the path towards supporting standardization and effective use of genomic data in the clinic.

Acknowledgements

1. Members of the HL7 Clinical Genomics Workgroup;
2. Members of the eMERGE EHRI Workgroup;
3. Steve Ordhal contributed through Microsoft's Partner Network that provides enhanced services to Johns Hopkins University.
4. Members of the BCM-HGSC Clinical Laboratory

Funding

This work was conducted under Phase III of the eMERGE Network, which was initiated and funded by the NHGRI through the following grants: U01HG008657 (Kaiser Permanente Washington/University of Washington); U01HG008685 (Brigham and Women's Hospital); U01HG008672 (Vanderbilt University Medical Center); U01HG008666 (Cincinnati Children's Hospital Medical Center); U01HG006379 (Mayo Clinic); U01HG008679 (Geisinger Clinic); U01HG008680 (Columbia University Health Sciences); U01HG008684 (Children's Hospital of Philadelphia); U01HG008673 (Northwestern University); U01HG008701 (Vanderbilt University Medical Center serving as the Coordinating Center); U01HG008676 (Partners Healthcare/Broad Institute); U01HG008664 (Baylor College of Medicine); and U54MD007593 (Meharry Medical College).

References

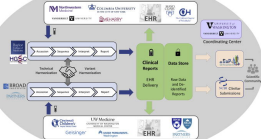
- [1] T.A. Manolio, R.L. Chisholm, B. Ozenberger, D.M. Roden, M.S. Williams, R. Wilson, D. Bick, E.P. Bottinger, M.H. Brilliant, C. Eng, K.A. Frazer, B. Korf, D.H. Ledbetter, J.R. Lupski, C. Marsh, D. Mrazek, M.F. Murray, P.H. O'Donnell, D.J. Rader, M.V. Relling, A.R. Shuldiner, D. Valle, R. Weinshilboum, E.D. Green, G.S. Ginsburg, Implementing genomic medicine in the clinic: the future is here, *Genet. Med. Off. J. Am. Coll. Med. Genet.* 15 (2013) 258–267. <https://doi.org/10.1038/gim.2012.157>.
- [2] T.A. Manolio, E.D. Green, Leading the way to genomic medicine, *Am. J. Med. Genet. C Semin. Med. Genet.* 166 (2014) 1–7. <https://doi.org/10.1002/ajmg.c.31384>.
- [3] E.D. Green, C. Gunter, L.G. Biesecker, V. Di Francesco, C.L. Easter, E.A. Feingold, A.L. Felsenfeld, D.J. Kaufman, E.A. Ostrander, W.J. Pavan, A.M. Phillippy, A.L. Wise, J.G. Dayal, B.J. Kish, A. Mandich, C.R. Wellington, K.A. Wetterstrand, S.A. Bates, D. Leja, S. Vasquez, W.A. Gahl, B.J. Graham, D.L. Kastner, P. Liu, L.L. Rodriguez, B.D. Solomon, V.L. Bonham, L.C. Brody, C.M. Hutter, T.A. Manolio, Strategic vision for improving human health at The Forefront of Genomics, *Nature.* 586 (2020) 683–692. <https://doi.org/10.1038/s41586-020-2817-4>.
- [4] Y. Yang, D.M. Muzny, F. Xia, Z. Niu, R. Person, Y. Ding, P. Ward, A. Braxton, M. Wang, C. Buhay, N. Veeraraghavan, A. Hawes, T. Chiang, M. Leduc, J. Beuten, J. Zhang, W. He, J. Scull, A. Willis, M. Landsverk, W.J. Craigen, M.R. Bekheirnia, A. Stray-Pedersen, P. Liu, S. Wen, W. Alcaraz, H. Cui, M. Walkiewicz, J. Reid, M. Bainbridge, A. Patel, E. Boerwinkle, A.L. Beaudet, J.R. Lupski, S.E. Plon, R.A. Gibbs, C.M. Eng, Molecular findings among patients referred for clinical whole-exome sequencing, *JAMA.* 312 (2014) 1870–1879. <https://doi.org/10.1001/jama.2014.14601>.
- [5] B.H. Shirts, J.S. Salama, S.J. Aronson, W.K. Chung, S.W. Gray, L.A. Hindorff, G.P. Jarvik, S.E. Plon, E.M. Stoffel, P.Z. Tarczy-Hornoch, E.M. Van Allen, K.E. Weck, C.G. Chute, R.R. Freimuth, R.W. Grundmeier, A.L. Hartzler, R. Li, P.L. Peissig, J.F. Peterson, L.V. Rasmussen, J.B. Starren, M.S. Williams, C.L. Overby, CSER and eMERGE: current and potential state of the display of genetic information in the electronic health record, *J. Am.*

- Med. Inform. Assoc. JAMIA. 22 (2015) 1231–1242. <https://doi.org/10.1093/jamia/ocv065>.
- [6] M.S. Williams, C.O. Taylor, N.A. Walton, S.R. Goehring, S. Aronson, R.R. Freimuth, L.V. Rasmussen, E.S. Hall, C.A. Prows, W.K. Chung, A. Fedotov, J. Nestor, C. Weng, R.K. Rowley, G.L. Wiesner, G.P. Jarvik, G. Del Fiol, Genomic Information for Clinicians in the Electronic Health Record: Lessons Learned From the Clinical Genome Resource Project and the Electronic Medical Records and Genomics Network, *Front. Genet.* 10 (2019). <https://doi.org/10.3389/fgene.2019.01059>.
- [7] M. Watkins, K. Eilbeck, FHIR Lab Reports: using SMART on FHIR and CDS Hooks to increase the clinical utility of pharmacogenomic laboratory test results, *AMIA Summits Transl. Sci. Proc.* 2020 (2020) 683–692. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7233102/> (accessed October 22, 2020).
- [8] G. Alterovitz, J. Warner, P. Zhang, Y. Chen, M. Ullman-Cullere, D. Kreda, I.S. Kohane, SMART on FHIR Genomics: facilitating standardized clinico-genomic apps, *J. Am. Med. Inform. Assoc. JAMIA.* 22 (2015) 1173–1178. <https://doi.org/10.1093/jamia/ocv045>.
- [9] B. Ryu, S.-Y. Shin, R.-M. Baek, J.-W. Kim, E. Heo, I. Kang, J.S. Yang, S. Yoo, Clinical Genomic Sequencing Reports in Electronic Health Record Systems Based on International Standards: Implementation Study, *J. Med. Internet Res.* 22 (2020). <https://doi.org/10.2196/15040>.
- [10] Electronic Medical Records and Genomics (eMERGE) Network, *Genome.Gov.* (n.d.). <https://www.genome.gov/Funded-Programs-Projects/Electronic-Medical-Records-and-Genomics-Network-eMERGE> (accessed August 26, 2020).
- [11] Summary - FHIR v4.0.1, (n.d.). <https://www.hl7.org/fhir/summary.html> (accessed October 22, 2020).
- [12] G.L. Wiesner, A. Kulchak Rahm, P. Appelbaum, S. Aufox, S.T. Bland, C.L. Blout, K.D. Christensen, W.K. Chung, E.W. Clayton, R.C. Green, M.H. Harr, N. Henrikson, C. Hoell, I.A. Holm, G.P. Jarvik, I.J. Kullo, P.E. Lammers, E.B. Larson, N.M. Lindor, M. Marasa, M. F. Myers, J.F. Peterson, C.A. Prows, J.D. Ralston, H. Milo Rasouly, R.R. Sharp, M.E. Smith, S.L. Van Driest, J.L. Williams, M.S. Williams, J. Wynn, K.A. Leppig, Returning Results in the Genomic Era: Initial Experiences of the eMERGE Network, *J. Pers. Med.* 10 (2020). <https://doi.org/10.3390/jpm10020030>.
- [13] Lessons learned from the eMERGE Network: balancing genomics in discovery and practice, *Hum. Genet. Genomics Adv.* 2 (2021) 100018. <https://doi.org/10.1016/j.xhgg.2020.100018>.
- [14] eMERGE Consortium. Electronic address: agibbs@bcm.edu, eMERGE Consortium, Harmonizing Clinical Sequencing and Interpretation for the eMERGE III Network, *Am. J. Hum. Genet.* 105 (2019) 588–605. <https://doi.org/10.1016/j.ajhg.2019.07.018>.
- [15] S. Aronson, L. Babb, D. Ames, R.A. Gibbs, E. Venner, J.J. Connelly, K. Marsolo, C. Weng, M.S. Williams, A.L. Hartzler, W.H. Liang, J.D. Ralston, E.B. Devine, S. Murphy, C.G. Chute, P.J. Caraballo, I.J. Kullo, R.R. Freimuth, L.V. Rasmussen, F.H. Wehbe, J.F. Peterson, J.R. Robinson, K. Wiley, C. Overby Taylor, eMERGE Network EHRI Working Group, Empowering genomic medicine by establishing critical sequencing result data flows: the eMERGE example, *J. Am. Med. Inform. Assoc. JAMIA.* 25 (2018) 1375–1381. <https://doi.org/10.1093/jamia/ocy051>.
- [16] <http://hl7.org/fhir/uv/genomics-reporting/index.html>, (n.d.).
- [17] HL7 Standards Product Brief - HL7 Version 2 Product Suite | HL7 International, (n.d.). https://www.hl7.org/implement/standards/product_brief.cfm?product_id=185 (accessed October 30, 2020).
- [18] HL7® FHIR® Accelerator Program | HL7 International, (n.d.). <http://www.hl7.org/about/fhir-accelerator/index.cfm> (accessed October 22, 2020).
- [19] HL7, Cloud Providers Unite for Healthcare Interoperability, (n.d.). <http://blog.hl7.org/cloud-providers-unite-for-healthcare-interoperability-fhir> (accessed October 22, 2020).
- [20] <https://grants.nih.gov/grants/guide/notice-files/NOT-OD-20-146.html>, (n.d.).

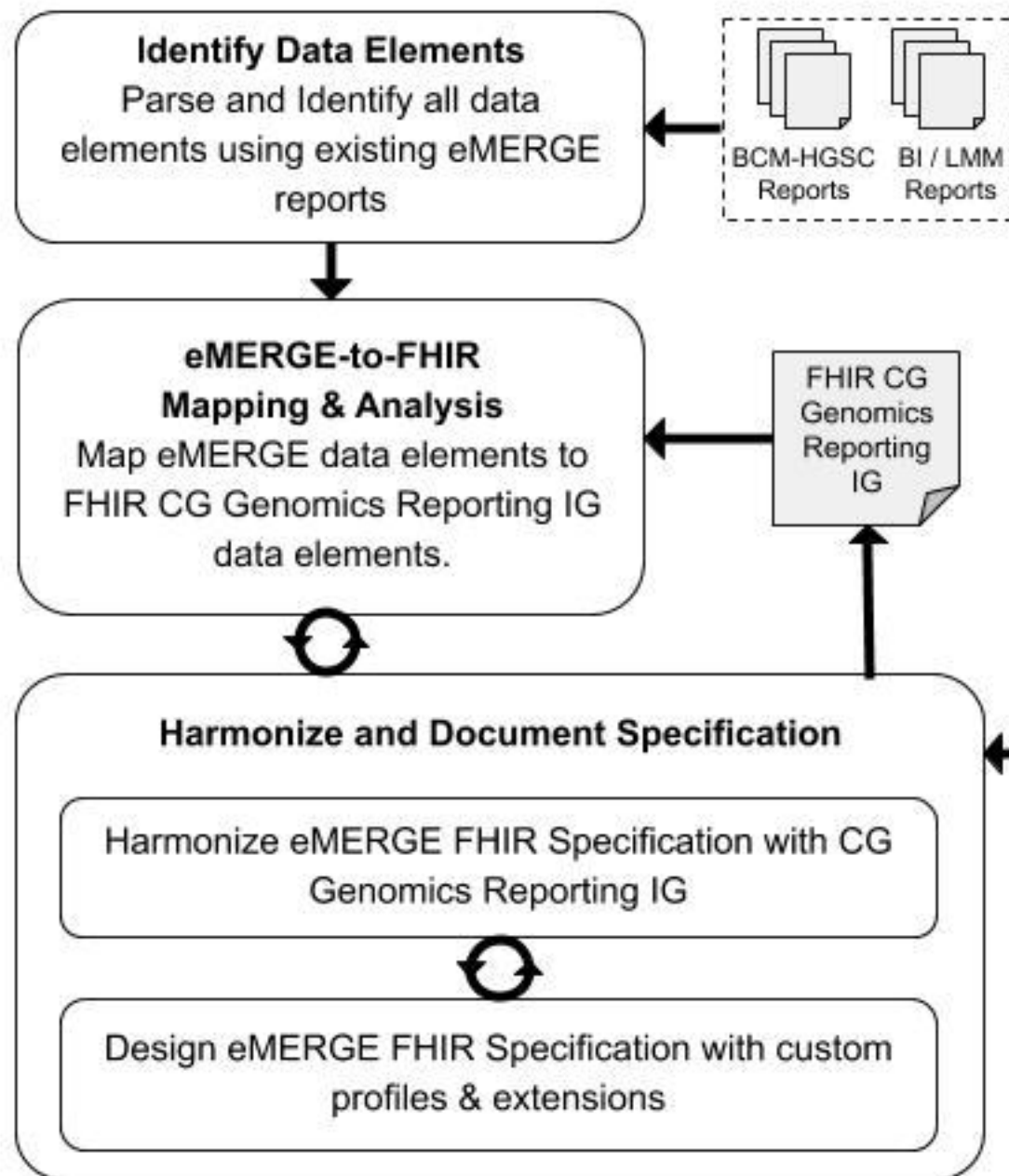
- [21] <https://grants.nih.gov/grants/guide/notice-files/NOT-OD-19-122.html>, (n.d.).
- [22] https://www.hl7.org/implement/standards/product_brief.cfm?product_id=491, (n.d.).
- [23] <https://chat.fhir.org/#narrow/stream/189875-genomics-.2F.20eMerge.20Pilot>, (n.d.).
- [24] <https://jira.hl7.org/projects/FHIR/issues/FHIR-29410?filter=allissues>, (n.d.).
- [25] emerge-ehri/FHIRGenomicsImplementation, eMERGE EHR Integration Workgroup, 2020. <https://github.com/emerge-ehri/FHIRGenomicsImplementation> (accessed September 26, 2020).
- [26] HAPI FHIR - The Open Source FHIR API for Java, (n.d.). <https://hapifhir.io/> (accessed October 24, 2020).
- [27] Smile CDR, Smile CDR Complete HL7 FHIR-Based Clin. Data Repos. (n.d.). <https://smilecdr.com/> (accessed October 24, 2020).
- [28] L.V. Rasmussen, M.E. Smith, F. Almaraz, S.D. Persell, L.J. Rasmussen-Torvik, J.A. Pacheco, R.L. Chisholm, C. Christensen, T.M. Herr, F.H. Wehbe, J.B. Starren, An ancillary genomics system to support the return of pharmacogenomic results, J. Am. Med. Inform. Assoc. 26 (2019) 306–310. <https://doi.org/10.1093/jamia/ocy187>.
- [29] welcome to eCDS_KB, (n.d.). <https://cdskb.org/> (accessed October 24, 2020).
- [30] emerge-ehri/fhir-server, GitHub. (n.d.). <https://github.com/emerge-ehri/fhir-server> (accessed September 26, 2020).
- [31] emerge-ehri/FHIRGenomicsImporter, eMERGE EHR Integration Workgroup, 2020. <https://github.com/emerge-ehri/FHIRGenomicsImporter> (accessed September 26, 2020).
- [32] emerge-ehri/FHIRGenomicsProxy, eMERGE EHR Integration Workgroup, 2020. <https://github.com/emerge-ehri/FHIRGenomicsProxy> (accessed September 26, 2020).
- [33] Azure API for FHIR(r) | Microsoft Azure, (n.d.). <https://azure.microsoft.com/en-us/services/azure-api-for-fhir/> (accessed October 26, 2020).
- [34] Johns Hopkins University - API Portal, (n.d.). <https://api.jh.edu/> (accessed October 26, 2020).
- [35] open.epic.com: Explore By Interface Type, (n.d.). <https://open.epic.com/Interface/FHIR> (accessed November 15, 2020).
- [36] Issues & Resolutions — emerge-fhir-spec 6e81023 documentation, (n.d.). https://emerge-fhir-spec.readthedocs.io/en/1.0/issues_and_resolutions.html# (accessed October 4, 2020).
- [37] Issues & Resolutions — emerge-fhir-spec 6e81023 documentation, (n.d.). https://emerge-fhir-spec.readthedocs.io/en/1.0/issues_and_resolutions.html#composite-reporting (accessed October 29, 2020).
- [38] Genomics Reporting Implementation Guide, (n.d.). <http://hl7.org/fhir/uv/genomics-reporting/grouper.html> (accessed October 29, 2020).
- [39] (44186) FHIR representation of a genetics test with multiple test... - FHIR Community - Zulip, (n.d.). <https://chat.fhir.org/#narrow/stream/189875-genomics-.2F.20eMerge.20Pilot/topic/FHIR.20representation.20of.20a.20genetics.20test.20with.20multiple.20test.2E.2E.2E> (accessed October 29, 2020).
- [40] eMERGE Results FHIR Specification — emerge-fhir-spec 6e81023 documentation, (n.d.). <https://emerge-fhir-spec.readthedocs.io/en/1.0/index.html> (accessed September 25, 2020).
- [41] Catalogue — emerge-fhir-spec 6e81023 documentation, (n.d.). <https://emerge-fhir-spec.readthedocs.io/en/1.0/artifacts/index.html#> (accessed September 17, 2020).
- [42] Extensions — emerge-fhir-spec 6e81023 documentation, (n.d.). <https://emerge-fhir-spec.readthedocs.io/en/1.0/artifacts/extensions.html> (accessed September 18, 2020).
- [43] emerge-ehri/fhir-specification, GitHub. (n.d.). <https://github.com/emerge-ehri/fhir-specification> (accessed September 26, 2020).
- [44] <https://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/index.html#safeharboriguidance>, (n.d.).
- [45] N.R.C. (US) C. on A.F. for D. a N.T. of Disease, eMERGE Consortium Data Use Agreement, National Academies Press (US), 2011.

- <https://www.ncbi.nlm.nih.gov/books/NBK92142/> (accessed January 26, 2021).
- [46] Sync for Genes | HealthIT.gov, (n.d.). <https://www.healthit.gov/topic/sync-genes> (accessed October 4, 2020).
- [47] R.H. Dolin, A. Boxwala, J. Shalaby, A Pharmacogenomics Clinical Decision Support Service Based on FHIR and CDS Hooks, *Methods Inf. Med.* 57 (2018) e115–e123. <https://doi.org/10.1055/s-0038-1676466>.
- [48] HLA-REPORTING\Home - FHIR v4.0.1, (n.d.). <http://fhir.b12x.org/ig/hla-reporting/> (accessed October 16, 2020).
- [49] G. Alterovitz, B. Heale, J. Jones, D. Kreda, F. Lin, L. Liu, X. Liu, K.D. Mandl, D.W. Poloway, R. Ramoni, A. Wagner, J.L. Warner, FHIR Genomics: enabling standardization for precision medicine use cases, *Npj Genomic Med.* 5 (2020) 1–4. <https://doi.org/10.1038/s41525-020-0115-6>.
- [50] HL7.FHIR.US.MCODE\Home - FHIR v4.0.1, (n.d.). <http://build.fhir.org/ig/HL7/fhir-mCODE-ig/branches/master/index.html> (accessed October 16, 2020).
- [51] <https://varnomen.hgvs.org/>, HGVS. (n.d.).
- [52] A. Simons, L.G. Shaffer, R.J. Hastings, Cytogenetic Nomenclature: Changes in the ISCN 2013 Compared to the 2009 Edition, *Cytogenet. Genome Res.* 141 (2013) 1–6. <https://doi.org/10.1159/000353118>.
- [53] P. Danecek, A. Auton, G. Abecasis, C.A. Albers, E. Banks, M.A. DePristo, R.E. Handsaker, G. Lunter, G.T. Marth, S.T. Sherry, G. McVean, R. Durbin, The variant call format and VCFtools, *Bioinformatics.* 27 (2011) 2156–2158. <https://doi.org/10.1093/bioinformatics/btr330>.
- [54] Home - SNP - NCBI, (n.d.). <https://www.ncbi.nlm.nih.gov/snp/> (accessed December 4, 2020).
- [55] ClinVar, (n.d.). <https://www.ncbi.nlm.nih.gov/clinvar/> (accessed December 4, 2020).
- [56] P. Pawliczek, R.Y. Patel, L.R. Ashmore, A.R. Jackson, C. Bizon, T. Nelson, B. Powell, R.R. Freimuth, N. Strande, N. Shah, S. Paithankar, M.W. Wright, S. Dwight, J. Zhen, M. Landrum, P. McGarvey, L. Babb, S.E. Plon, A. Milosavljevic, on behalf of the Clinical Genome (ClinGen) Resource, ClinGen Allele Registry links information about genetic variants, *Hum. Mutat.* 39 (2018) 1690–1701. <https://doi.org/10.1002/humu.23637>.
- [57] Cosmic, COSMIC - Catalogue of Somatic Mutations in Cancer, (n.d.). <https://cancer.sanger.ac.uk/cosmic> (accessed December 4, 2020).
- [58] GA4GH Variation Representation Specification — GA4GH Variation Representation Specification HEAD documentation, (n.d.). <https://vr-spec.readthedocs.io/en/1.1/> (accessed November 18, 2020).
- [59] The GA4GH Variation Representation Specification (VRS): a Computational Framework for the Precise Representation and Federated Identification of Molecular Variation | bioRxiv, (n.d.). <https://www.biorxiv.org/content/10.1101/2021.01.15.426843v1> (accessed January 26, 2021).

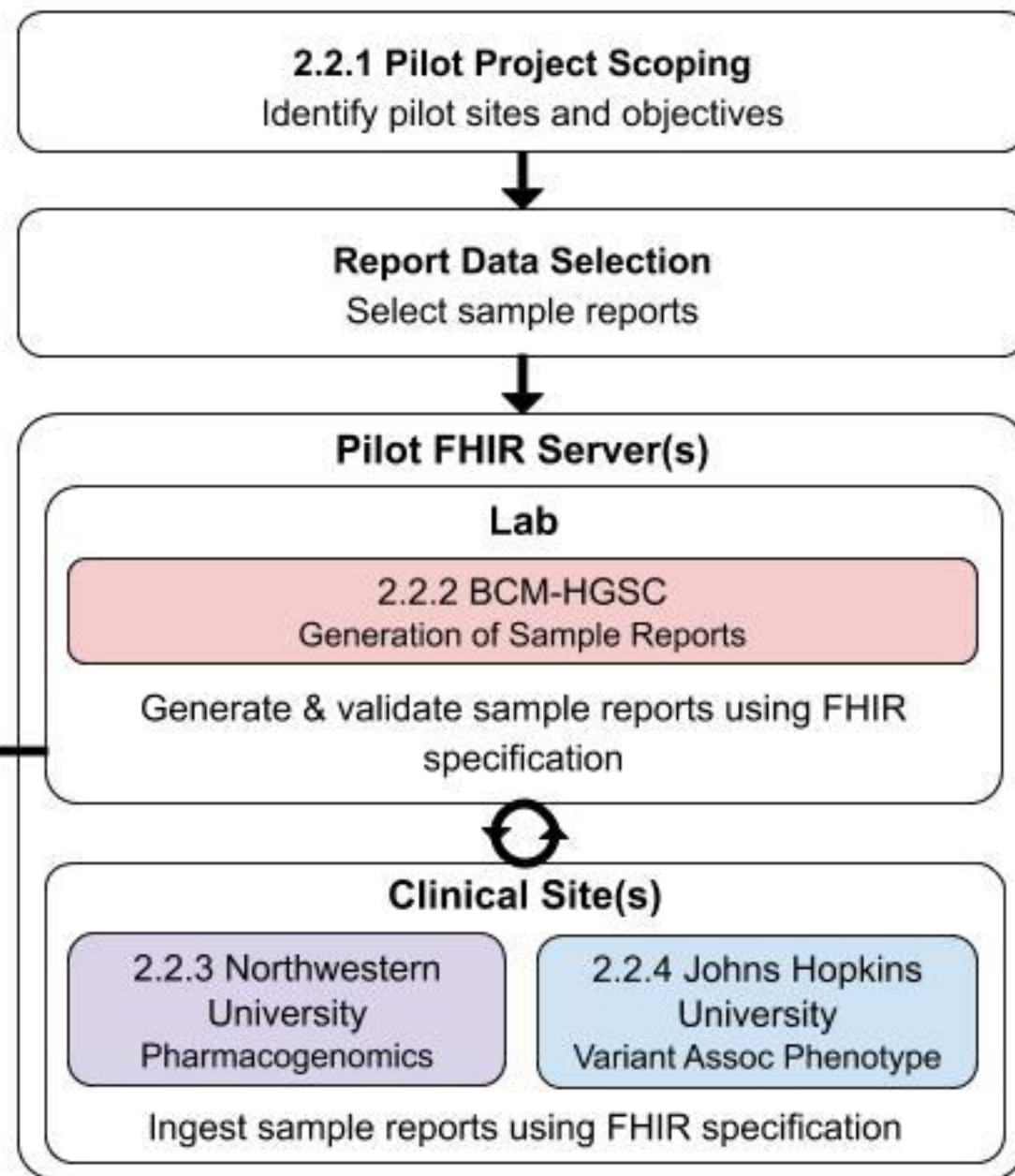
eMERGE III Clinical and Discovery Platform



2.1 eMERGE FHIR Specification Development



2.2 Use Case Pilot Projects Development





Patient Name:		Sample Collected Date:	09/29/2016
Patient ID:	XXXXXXXXXX	Sample Received Date:	02/22/2018
Age:	x	Report Date:	xx/xx/yyyy
DOB:	xxxxx	Sample Type:	DNA
Sex:	Male	Indication for Testing:	Not selected for trait
Patient Sample ID:	XXXXXXXXXXXXX	Ordering Physician Name:	Freda Feelgood 3615 Civic Center Blvd. ...
Accession #:	XXXXXXXXXXXXXX		

eMERGE-Seq Panel Sequencing Report
Everywhere Hospital

This test interrogates the protein-coding and exon-splicing regions of 109 genes as well as 1551 single-nucleotide polymorphisms that may impact human health and disease. Clinical interpretation and reporting are provided for pathogenic and likely pathogenic variants for genes and single nucleotide polymorphisms as described in the methodology section.

PATHOGENIC AND/OR LIKELY PATHOGENIC VARIANTS DETECTED

A heterozygous c.1552C>T (p.Arg518*) pathogenic variant in the KCNQ1 (NM_000218.2) gene was detected in this individual, which was confirmed by Sanger sequencing. Defects in KCNQ1 are the cause of long QT syndrome 1 (LQT1) [MIM 192500], an autosomal dominant heart disorder characterized by a prolonged QT interval on the ECG and polymorphic ventricular arrhythmias. They cause syncope and sudden death, a response to exercise or fright, and are associated with a potential event of a sudden cardiac death in infancy. The variant c.1552C>T (p.Arg518*) was reported in ClinVar (accession number VCV00022539601). Based on these data the p.Arg518* variant in the KCNQ1 gene is classified as pathogenic.

Table 1: Details of Pathogenic and Likely Pathogenic Variants

Disease	Inh.	Gene	Position (NCBI 37)	Variant	Zyg.	Notes	Interpretation
Long QT syndrome 1 [MIM 192500]; Jervell and Lange-Nielsen syndrome 1 [MIM 607542]	AD/AR	KCNQ1	chr11 g.2790111C>T	c.1552C>T p.Arg518*	Heterozygous	PMID 10482963, 10704188, 10737999, 11530100,18752142, 22539601, 24552659, 27451284, 28438721; rs17215500 Confirmed by Sanger sequencing	Pathogenic

Table 2: Details of Copy Number Variants:

None detected

Table 3: Details of Pharmacogenomic Variants

Pharmacogenomics variants are returned for the following genes: CYP2C19, DPYD, INFL3, SLCO1B1, TPMT, CYP2C9/VKORC1. Star alleles are determined based on the variants detected by this assay. Star alleles may not be accurately defined due to the limitations of this assay which include: 1) The presence of additional variants defining functional and non functional alleles in a patient, not detected by this assay, and 2) the lack of ability to determine the phase of the variants when a star allele is defined by multiple variants. Additionally, undetected genetic and/or non genetic factors such as drug-drug interactions, may also impact the phenotype. This pharmacogenomic report is limited to CPIC level A alleles and drug recommendations. Additional (level B and lower) drugs may be metabolized by these reported enzymes; and additional enzymes, not reported here, may affect the metabolism of a reported drug. Refer to the current recommendation for dosage guidelines. See Methodology for details.

Gene	Drug	Diplotype	Phenotype	Recommendation
CYP2C19	clopidogrel	*2/*2	Poor metabolizer	https://cpicpgx.org/guidelines/guideline-for-clopidogrel-and-cyp2c19/
	voriconazole			https://cpicpgx.org/guidelines/guideline-for-voriconazole-and-cyp2c19/
	citalopram, escitalopram			https://cpicpgx.org/guidelines/guideline-for-selective-serotonin-reuptake-inhibitors-and-cyp2d6-and-cyp2c19/
	amitriptylin			https://cpicpgx.org/guidelines/guideline-for-tricyclic-antidepressants-and-cyp2d6-and-cyp2c19/
DPYD	capecitabine, fluorouracil, tegafur	*1/*1	Normal DPD activity and "normal" risk for fluoropyrimidine toxicity	https://cpicpgx.org/guidelines/guideline-for-fluoropyrimidines-and-dpyd/
IFNL3	peginterferon alfa-2a, peginterferon alfa-2b, ribavirin	rs12979860 C/C	Favorable response genotype	https://cpicpgx.org/guidelines/guideline-for-peg-interferon-alpha-based-regimens-and-ifnl3/
SLCO1B1	simvastatin	rs4149056 T/T	Normal function, Normal simvastatin induced myopathy risk	https://cpicpgx.org/guidelines/guideline-for-simvastatin-and-slco1b1/
TPMT	azathioprine, mercaptopurine, thioguanine	*1/*1	High activity	https://cpicpgx.org/guidelines/guideline-for-thiopurines-and-tpmt/ thioguanine
CYP2C9	warfarin	*1/*1	Extensive metabolizer	https://cpicpgx.org/guidelines/guideline-for-warfarin-and-cyp2c9-and-vkorc1/
VKORC1		C/T		

Interpretation of Pharmacogenomic Variants:

This individual is homozygous for a non-functional allele of the CYP2C19 gene. Based on the genotype result, this patient is predicted to have a CYP2C19 poor metabolizer phenotype. This genotype information can be used by patients and clinicians as part of the shared decision-making process for several drugs metabolized by CYP2C19 including clopidogrel, voriconazole, amitriptyline, citalopram and escitalopram. For clopidogrel, individuals with this diplotype are expected to have significantly reduced platelet inhibition, increased residual normal CYP2C9 function.This individual is also heterozygous for the variant allele for the VKORC1 gene. Expression level of the VKORC1 gene is associated with warfarin sensitivity. Based on the genotype result, this patient is predicted to have medium sensitivity to warfarin.

Comments & Recommendations:

It is recommended that correlation of these findings with the clinical phenotype be performed. Genetic counseling for the patient and at-risk family members is recommended.

Gene Coverage:

All genes have 100% of targeted bases sequenced to redundant coverage of 20x or greater with the following exceptions: APOB (99.96%), CACNA1B (97.08%), COL5A1 (99.37%), FLG (99.97%), KCNQ1 (97.46%), RYR1 (98.64%), TGFBR1 (93.56%). Further information, including specific coverage for this patient's sample, is available in the ExCID report.

Methodology:

- eMERGE-Seq Version 2 NGS Panel: for the paired-end pre-capture library procedure, genome DNA is fragmented by sonicating genome DNA and ligating to the Illumina multiplexing PE adapters (reference 1). The adapter-ligated DNA is then PCR amplified using primers with sequencing barcodes (indexes). For target enrichment capture procedure, the pre-capture library is enriched by hybridizing to biotin labeled probes (reference 2).
- For the pharmacogenomic variants, the star alleles are determined based on the variants detected by this assay. Alleles reported for TPMT are limited to *1, *2, *3A, *3B, *3C and *4. Alleles reported for CYP2C19 are limited to *1, *2, *4A, *4B, *5, *6, *7, *8, *17. If reported, alleles for DPD are limited to *1, *2A, *13 and rs67376798. Alleles reported for CYP2C9 are limited to *1, *2 and *3; and rs9923231 for VKORC1. Additional rare star alleles have been reported with reduced or no function for TPMT, CYP2C19 and DPD; however, the variants defining these additional star alleles are not detected with this assay. For SLCO1B1, this assay only detects rs4149056. The minor C allele at rs4149056 defines the SLCO1B1*5 (rs4149056 alone) but also tags the *15 and *17 alleles. Thus a *5 allele may represent a *15 or *17 allele. However, the magnitude of the phenotypic effect is similar for *5, *15, and *17 alleles.

References:

- Illumina, Inc. (2011) Multiplexing Sample Preparation Guide (Part # 1005361 Rev. D). 2011.
- Roche NimbleGen, Inc. (2010) NimbleGen SeqCap EZ Exome Library SR User's Guide (Version 2.2).

Medicine (2015) 17, 405-423. PMID: 25741868.

Jane Doe Ph.D
ABMGG Certified Molecular Geneticist
Laboratory Director
09/30/2018

John Smith, M.D., FACMG
ABMGG Certified Molecular Geneticist
Medical Director
09/30/2018

This test was developed and its performance determined by this laboratory. It has not been cleared or approved by U.S. Food and Drug Administration. Since FDA is not required for clinical use of this test, this laboratory has established and validated the test's accuracy and precision, pursuant to the requirement of CLIA '88. This laboratory is licensed and/or accredited under CLIA and CAP (CAP# 9999999 / CLIA# 99D99999999).

Laboratory for Molecular Medicine
65 Landsdowne Street, Cambridge MA 02139
Phone: (999)999-9999 Fax: (888)888-8888
www.lab.org/personalizedmedicine

Lab Accession: PM-XX-XXXXXX
Patient Name: LAST, FIRST
Birth Date: 1/1/1800
Age Sex: 212 Year old Female

MOLECULAR DIAGNOSTICS REPORT

Specimen Type: DNA, Isolated - Blood, Peripheral
Related Accession(s):
Referring Physician: DOCTOR SMITH PHD
Copies To: XXXXX LAB

Received Date: 3/22/2017

Referring Facility: ABC CLINIC

Referring Fac. MRN:

Lab Control Number: YYYYYY-UUUUUU-XXXXXX

Family Number: F0xxxxxxx

TEST DESCRIPTION - Copy Number Variation Analysis
Sequence Confirmation Test
eMERGE III Sequencing Panel

TEST PERFORMED - CNV-a; SeqConV2; EMERGE-pnlE

INDICATION FOR TEST - Hyperlipidemia

RESULTS

DNA VARIANTS:

Heterozygous c.501C>A (p.Cys167X), Exon 4, LDLR, Pathogenic
Heterozygous c.148C>T (p.Arg50Cys), Exon 3, SDHC, Likely pathogenic

INTERPRETATION:

Positive. DNA sequencing of the coding regions and splice sites of a subset of 53 genes from the eMERGE III sequencing panel (see methodology section below) identified the variants listed above. Copy number analysis using NGS could not be completed because data did not meet quality standards for CNV detection. For a list of exons that are incompletely covered please see "Additional notes" section below.

SUMMARY (VARIANTS RELEVANT TO INDICATION FOR TESTING):

One heterozygous pathogenic variant in the LDLR gene was identified in this individual. Pathogenic variants in the LDLR gene cause familial hypercholesterolemia (FH), which is consistent with this individual's reported diagnosis. The available information is described in the detailed variant interpretation section.

FH due to pathogenic variants in the LDLR gene is typically inherited in an autosomal dominant pattern. Each first-degree relative has a 50% (or 1 in 2) chance of inheriting a variant and its risk for disease.

SUMMARY (OTHER VARIANTS OF MEDICAL SIGNIFICANCE):

One heterozygous likely pathogenic variant in the SDHC gene was identified in this individual. Based on the available evidence, this individual may be at risk for developing hereditary paraganglioma-pheochromocytoma syndrome. The available information is described in the detailed variant interpretation section.

Hereditary paraganglioma-pheochromocytoma syndrome due to pathogenic variants in the SDHC gene is typically inherited in an autosomal dominant pattern. Each first-degree relative has a 50% (or 1 in 2) chance of inheriting a variant and its risk for disease.

ADDITIONAL NOTES:

Disease penetrance and severity can vary due to modifier genes and/or environmental factors. The significance of a variant should therefore be interpreted in the context of the individual's clinical manifestations.

Variants of unlikely clinical significance are not included on eMERGE reports but can be shared separately. Please note that variant classifications may change over time if more information becomes available. Please contact us at 617-768-8500 or LMM@partners.org.

We cannot exclude the presence of a variant in the following genes(s) due to <95% coverage at 20X*: KCNQ1(94.7%), PTEN(92.9%), RB1(91.8%), TGFBR1(93.3%)

* Additional information (gene and variant lists as well as reference sequence) available at <http://personalizedmedicine.partners.org/Laboratory-For-Molecular-Medicine/>.

DETAILED VARIANT INTERPRETATIONS:

p.Cys167X, c.501C>A (LDLR; NM_000527.4; Chr19g.11216083C>A; GRCh37):

The p.Cys167X variant in LDLR (also described as p.Cys146X in the literature) has been reported in the heterozygous state in >10 individuals with familial hypercholesterolemia (FH), segregated with disease in one affected relative from one family (Lombardi 1995, Heath 1999, Fouchier 2001, Bodamer 2002, van der Graaf 2011, Tichy 2012, Sharifi 2016) and was absent from large population studies. Additionally, this variant has been reported by other clinical laboratories in ClinVar (Variation ID: 200918). This nonsense variant leads to a premature termination codon at position 167, which is predicted to lead to a truncated or absent protein. Heterozygous loss of function of the LDLR gene is an established disease mechanism in FH. In summary, this variant meets our criteria to be classified as pathogenic for familial hypercholesterolemia in an autosomal dominant manner based upon predicted impact to the protein, presence in multiple affected individuals and absence in the general population. ACMG/AMP Criteria applied (Richards 2015): PVS1, PS4. Moderate, PM2.

p.Arg50Cys, c.148C>T (SDHC; NM_003001.3; Chr1q.161298256C>T; GRCh37):

The p.Arg50Cys variant in SDHC has been reported in 4 individuals with hereditary paraganglioma-pheochromocytoma syndrome

required to fully establish its clinical significance, the p.Arg50Cys variant is likely pathogenic. ACMG/AMP Criteria applied (Richards 2015): PM2; PS4 Moderate; PP3; PS3 Supporting.

RECOMMENDATION:

RECOMMENDATION: Genetic counseling is recommended for this individual and their relatives. Familial variant testing is available for other relatives if desired. For assistance in locating genetic counseling services or disease specialists, please call the laboratory at 617-768-8500 or email at LMM@partners.org.

TEST INFORMATION

BACKGROUND:

The eMERGE (electronic Medical Records and GENomics) network combines DNA biorepositories with electronic health record (EHR) systems for large-scale discovery and clinical implementation research in genomic medicine. A main goal is the return of genomic testing results to patients in a clinical care setting. In phase III, participating sites are sequencing 109 clinically relevant genes in ~25,000 participants using a custom next generation sequencing panel.

METHODOLOGY:

Test content (target region): This test includes 109 genes (including the ACMG56 genes; PMID: 23788249, and additional genomic positions for known variants. For reference sequences exons/positions covered see <http://personalizedmedicine.partners.org/Laboratory-For-Molecular-Medicine/>). Note that this test may not detect variants in regions with difficult sequence contexts (e.g. high or low GC content) and is generally not designed to detect deep intronic variants as well as variants in the 5' and 3'UTR. Regions with high sequence homology are only included in this test if a unique Sanger sequencing assay can be designed to rule out false positive calls.

been cleared or approved by the U.S. Food and Drug Administration (FDA). The FDA has determined that such clearance or approval is not necessary.

REFERENCES:

Bennedbaek M, Rossing M, Rasmussen ÅK, Gerdes AM, Skytte AB, Jensen UB, Nielsen FC, Hansen TVO. 2016. Identification of eight novel SDHB, SDHC, SDHD germline variants in Danish pheochromocytoma/paraganglioma patients. *Hered Cancer Clin Pract*. 14:13.

Bodamer OA, Bercovich D, Schlabach M, Ballantyne C, Zoch D, Beaudet AL. 2002. Use of denaturing HPLC to provide efficient detection of mutations causing familial hypercholesterolemia. Clin. Chem. 48(11):1913-8.

Bodian DL, McCutcheon JN, Kothiyal P, Huddleston KC, Iyer RK, Vockley JG, Niederhuber JE. 2014. Germline variation in cancer-susceptibility genes in a healthy, ancestrally diverse cohort: implications for individual genome sequencing. PLoS ONE. 9(4):e94554.

Wiegman A. 2011. Molecular basis of autosomal dominant hypercholesterolemia: assessment in a large cohort of hypercholesterolemic children. *Circulation*. 123(11):1167-73.

REPORT PREPARATION by Jane Smith Ph.D., on Tuesday February 20, 2018 at 02:43:27PM
REPORT by Hana Helpful Ph.D., on Wednesday February 21, 2018 at 12:28:09PM
Final Diagnosis by **Hana Helpful Ph.D.**, Electronically signed on Wednesday February 21, 2018 at 12:29:47PM

DOCTOR SMITH, MD, PhD
ABC CLINIC
MAIN STREET
ANYTOWN, USA 99999-9999

Donna Director, Ph.D., FACMG, Director
www.lab.org/personalizedmedicine
CLIA# : 88D888888888

Report

1

Header

Performing Lab

7

Case

Patient

Sample/Specimen

Request / Orderer

Test Performed Name

Body

Diagnostic Gene Panel Results Group

Overall Interpretation

Test Background

Clinical
Interpretation

Identified Variant
Genotype

PGx Gene Panel Results Group

Test Background

Medication
Implication

Identified Variant
Genotype/Diplotype

Recommendations

Comments

Gene Coverage

Test Performed Methodology/References

Footer

Results Interpreter(s)

Test Disclaimer

Patient Name:		Sample Collected Date:	09/29/2016
Patient ID:	XXXXXXXXXX	Sample Received Date:	02/22/2018
Age:	x	Report Date:	xx/xx/yyyy
DOB:	xxxxx	Sample Type:	DNA
Sex:	Male	Indication for Testing:	Not selected for trait
Patient Sample ID:	XXXXXXXXXXXXXX	Ordering Physician Name:	Freda Feelgood 3615 Civic Center Blvd. ...
Accession #:	XXXXXXXXXXXXXX		

eMERGE-Seq Panel Sequencing Report

Everywhere Hospital

This test interrogates the protein-coding and exon-splicing regions of 109 genes as well as 1551 single-nucleotide polymorphisms that may impact human health and disease. Clinical interpretation and reporting are provided for pathogenic and likely pathogenic variants for genes and single nucleotide polymorphisms as described in the methodology section.

PATHOGENIC AND/OR LIKELY PATHOGENIC VARIANTS DETECTED

A heterozygous c.1552C>T (p.Arg518*) pathogenic variant in the KCNQ1 (NM_000218.2) gene was detected in this individual, which was confirmed by Sanger sequencing. Defects in KCNQ1 are the cause of long QT syndrome 1 (LQT1) [MIM 192500], an autosomal dominant heart disorder characterized by a prolonged QT interval on the ECG and polymorphic ventricular arrhythmias. They cause syncope and

Table 1: Details of Pathogenic and Likely Pathogenic Variants

Disease	Inh.	Gene	Position (NCBI 37)	Variant	Zyg.	Notes	Interpretation
Long QT syndrome 1 [MIM 192500]; Jervell and Lange-Nielsen syndrome 1AD/AR [MIM 607542]		KCNQ1	chr11 g.2790111C>T	c.1552C>T p.Arg518*	Heterozygous	PMID 10482963, 10704188, 10737999, 11530100,18752142, 22539601, 24552659, 27451284, 28438721; rs17215500 Confirmed by Sanger sequencing	Pathogenic

Table 2: Details of Copy Number Variants:

None detected

Table 3: Details of Pharmacogenomic Variants

Pharmacogenomics variants are returned for the following genes: CYP2C19, DPYD, INFL3, SLCO1B1, TPMT, CYP2C9/VKORC1. Star alleles are determined based on the variants detected by this assay. Star alleles may not be accurately defined due to the limitations of this assay which include: 1) The presence of additional variants defining functional and non functional alleles in a patient, not detected by this assay, and 2) the lack of ability to determine the phase of the variants when a star allele is defined by multiple variants. Additionally, undetected genetic and/or non genetic factors such as drug-drug interactions, may also impact the phenotype. This pharmacogenomic report is limited to CPIC level A alleles and drug recommendations. Additional (level B and lower) drugs may be metabolized by these reported enzymes; and additional enzymes, not reported here, may affect the metabolism of a reported drug. Refer to the current recommendation for dosage guidelines. See Methodology for details.

Gene	Drug	Diplotype	Phenotype	Recommendation
CYP2C19	clopidogrel	*2/*2	Poor metabolizer	https://cpicpgx.org/guidelines/guideline-for-clopidogrel-and-cyp2c19/
	voriconazole			https://cpicpgx.org/guidelines/guideline-for-voriconazole-and-cyp2c19/
	citalopram, escitalopram			https://cpicpgx.org/guidelines/guideline-for-selective-serotonin-reuptake-inhibitors-and-cyp2d6-and-cyp2c19/
	amitriptylin			https://cpicpgx.org/guidelines/guideline-for-tricyclic-antidepressants-and-cyp2d6-and-cyp2c19/
DPYD	capecitabine, fluorouracil, tegafur	*1/*1	Normal DPD activity and "normal" risk for fluoropyrimidine toxicity	https://cpicpgx.org/guidelines/guideline-for-fluoropyrimidines-and-dpyd/
IFNL3	peginterferon alfa-2a, peginterferon alfa-2b, ribavirin	rs12979860 C/C	Favorable response genotype	https://cpicpgx.org/guidelines/guideline-for-peg-interferon-alpha-based-regimens-and-ifnl3/
SLCO1B1	simvastatin	rs4149056 T/T	Normal function, Normal simvastatin induced myopathy risk	https://cpicpgx.org/guidelines/guideline-for-simvastatin-and-slco1b1/
TPMT	azathioprine, mercaptopurine, thioguanine	*1/*1	High activity	https://cpicpgx.org/guidelines/guideline-for-thiopurines-and-tpmt/ thioguanine
CYP2C9	warfarin	*1/*1	Extensive metabolizer	https://cpicpgx.org/guidelines/guideline-for-warfarin-and-cyp2c9-and-vkorc1/
VKORC1		C/T		

Interpretation of Pharmacogenomic Variants:

This individual is homozygous for a non-functional allele of the CYP2C19 gene. Based on the genotype result, this patient is predicted to have a CYP2C19 poor metabolizer phenotype. This genotype information can be used by patients and clinicians as part of the shared decision-making process for several drugs metabolized by CYP2C19 including clopidogrel, voriconazole, amitriptyline, citalopram and escitalopram. For clopidogrel, individuals with this diplotype are expected to have significantly reduced platelet inhibition, increased residual

This individual is homozygous for the normal allele for the CYP2C9 gene. Based on the genotype result, this patient is predicted to have normal CYP2C9 function. This individual is also heterozygous for the variant allele for the VKORC1 gene. Expression level of the VKORC1 gene is associated with warfarin sensitivity. Based on the genotype result, this patient is predicted to have medium sensitivity to warfarin.

Comments & Recommendations:

It is recommended that correlation of these findings with the clinical phenotype be performed. Genetic counseling for the patient and at-risk family members is recommended.

Gene Coverage:

All genes have 100% of targeted bases sequenced to redundant coverage of 20x or greater with the following exceptions: APOB (99.96%), CACNA1B (97.08%), COL5A1 (99.37%), FLG (99.97%), KCNQ1 (97.46%), RYR1 (98.64%), TGFBR1 (93.56%). Further information, including specific coverage for this patient's sample, is available in the ExCID report.

Methodology:

1. eMERGE-Seq Version 2 NGS Panel: for the paired-end pre-capture library procedure, genome DNA is fragmented by sonicating genome DNA and ligating to the Illumina multiplexing PE adapters (reference 1). The adapter-ligated DNA is then PCR amplified using primers with sequencing barcodes (indexes). For target enrichment capture procedure, the pre-capture library is enriched by hybridizing to biotin labeled in-solution probes (reference 2) at 56°C for 16 - 19 hours. For massively parallel sequencing, the post-capture library DNA is subjected to

6. For the pharmacogenomic variants, the star alleles are determined based on the variants detected by this assay. Alleles reported for TPMT are limited to *1, *2, *3A, *3B, *3C and *4. Alleles reported for CYP2C19 are limited to *1, *2, *4A, *4B, *5, *6, *7, *8, *17. If reported, alleles for DPD are limited to *1, *2A, *13 and rs67376798. Alleles reported for CYP2C9 are limited to *1, *2 and *3; and rs9923231 for VKORC1. Additional rare star alleles have been reported with reduced or no function for TPMT, CYP2C19 and DPD; however, the variants defining these additional star alleles are not detected with this assay. For SLCO1B1, this assay only detects rs4149056. The minor C allele at rs4149056 defines the SLCO1B1*5 (rs4149056 alone) but also tags the *15 and *17 alleles. Thus a *5 allele may represent a *15 or *17 allele. However, the magnitude of the phenotypic effect is similar for *5, *15, and *17 alleles.

References:

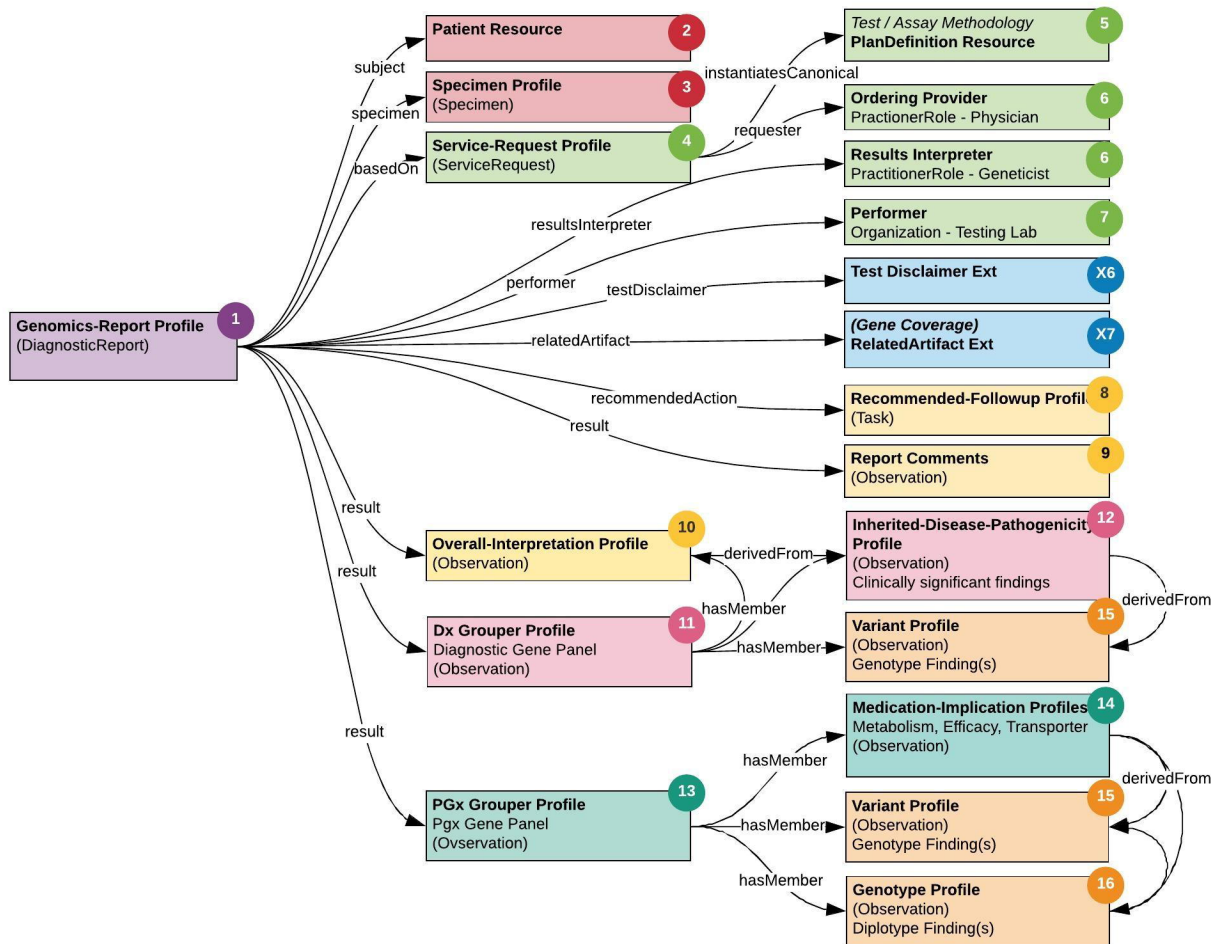
1. Illumina, Inc. (2011) Multiplexing Sample Preparation Guide (Part # 1005361 Rev. D). 2011.
2. Roche NimbleGen, Inc. (2010) NimbleGen SeqCap EZ Exome Library SR User's Guide (Version 2.2).

Medicine (2015) 17, 405-423. PMID: 25741868.

Jane Doe Ph.D
ABMGG Certified Molecular Geneticist
Laboratory Director
09/30/2018

John Smith, M.D., FACMG
ABMG Certified Molecular Geneticist
Medical Director
09/30/2018

This test was developed and its performance determined by this laboratory. It has not been cleared or approved by U.S. Food and Drug Administration. Since FDA is not required for clinical use of this test, this laboratory has established and validated the test's accuracy and precision, pursuant to the requirement of CLIA '88. This laboratory is licensed and/or accredited under CLIA and CAP (CAP# 9999999 / CLIA# 99D999999999).



eMERGE III FHIR Pilot

eMERGE III

Accession
Sequence
Interpret
Report

3.2.1 BCM-HGSC - Generation of Sample Reports

1. Identity Pilot Reports

3. Store FHIR Reports in S3



2. Generate FHIR Report Bundles



4.a Pull from BCM S3

3.2.2 Northwestern University - Pharmacogenomics



5.a. Validate, Store & Load in AGS



6.a. HL7 V2 to EHR for PGx CDS

4.b Push to JHU Server

3.2.3 Johns Hopkins University - Variant Associated Phenotype



5.b. Validate, Store & Push to EHR



6.b. Validate & add to Patient Problem List



Standard Reporting Use Cases	
Diagnostic Gene Panel	
Reported Variants	Single-nucleotide variants (SNVs), Insertions or deletions (InDels), copy number variants (CNVs)
Medically Significant Findings for Indication	Pathogenic, Likely Pathogenic, Uncertain Significance
Secondary Findings	Medically Significant Findings in a gene unrelated to the indication for testing
Overall Interpretations	Positive, Negative, Inconclusive
Pharmacogenomic Gene Panel	
PGx Phenotype Findings	Efficacy, Metabolism, Transporter Function
Reported Diplotypes	Simple, Complex, Wildtype, Non-Wild Type
Specialized Reporting Use Cases	
LDLR Scores-Polygenic Risk Scores	Mayo Clinic, an eMERGE Study Site
Research Only Reports	Columbia University, an eMERGE Study Site

Issue	Cat.	Resolution
#1 Composite Reporting	Major	GR IG Change
#2 Nested Results	Major	Clarification
#3 Test Information	Major	Clarification
#4 Interpretation Summary	Major	Workaround
#5 Gene/Region Coverage	Major	Workaround
#6 Recommendations	Major	Clarification
#7 Secondary Findings	Major	Clarification
#8 Variant Data Types	Major	Deferred
#9 Report Comments	Minor	Clarification
#10 Confirmation Testing	Major	Workaround
#11 Pathogenicity Phenotypes	Minor	GR IG Change

Issue	Cat.	Resolution
#12 Pathogenicity Values	Minor	GR IG Change
#13 Report Category	Minor	GR IG Change
#14 Assessed Med Citations	Major	Workaround
#15 Sign-Out v Sent Date	Minor	Workaround
#16 Recommended Followup Reasons	Minor	Workaround
#17 Recommended Followup Codes	Minor	GR IG Change
#18 Disclaimers	Major	Workaround
#19 Patient Internal ID	Minor	Clarification
#20 Patient Age	Major	Workaround
#21 Research Flag	Minor	Deferred

Resources	Profile or Resource	Elements				
		Total	Ref	Ext	Id/Code	Other
GenomicsReport	profile	22	9	3	6	4
Report Resources						
Patient	resource	10	0	4	3	3
Specimen	profile	11	2	0	5	4
RecommendedFollowup	profile	10	2	0	5	3
ServiceRequest	profile	17	5	0	9	3
PlanDefinition	resource	11	0	0	4	7
Observation Resources						
OverallInterpretation	profile	15	5	1	6	3
ReportComment	resource	12	4	0	5	3
GrouperDx	profile	16	7	1	5	3
GrouperPgx	profile	17	9	0	5	3
InheritedDiseasePathogenicity	profile	19	5	1	9	4
MedicationImplication*	profile	17	6	1	7	3
Variant	profile	31	4	0	21	6
Genotype	profile	15	5	0	7	3
Other Resources						
PractitionerRole	resource	6	2	0	3	1
Practitioner	resource	7	0	0	4	3
Organization	resource	9	1	0	4	4
Total Concept Elements		223	63	11	97	52